

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2000, 21:41:12 ; Search time 1895.15 Seconds
(without alignments)
-951.699 Million cell updates/sec

Title: US-08-988-242-1_COPY(1232_1825)
Perfect score: 594
Sequence: 1 CAGGTACAGCTAACGGCTT.....AAACGAGCTCCGTGGCGATA 594

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0
Searched: 821193 seqs, -1518192014 residues
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_bal:*
 - 2: gb_ba2:*
 - 3: gb_on:*
 - 4: gb_ov:*
 - 5: gb_pat:*
 - 6: gb_ph:*
 - 7: gb_pl1:*
 - 8: gb_pl2:*
 - 9: gb_pr1:*
 - 10: gb_pr2:*
 - 11: gb_pr3:*
 - 12: gb_ro:*
 - 13: gb_sts:*
 - 14: gb_sy:*
 - 15: gb_un:*
 - 16: gb_vt:*
 - 17: em_fun:*
 - 18: em_hum1:*
 - 19: em_hum2:*
 - 20: em_in:*
 - 21: em_on:*
 - 22: em_or:*
 - 23: em_ov:*
 - 24: em_pat:*
 - 25: em_ph:*
 - 26: em_pl:*
 - 27: em_ro:*
 - 28: em_sts:*
 - 29: em_sy:*
 - 30: em_un:*
 - 31: em_vt:*
 - 32: gb_htg1:*
 - 33: gb_htg2:*
 - 34: gb_in1:*
 - 35: gb_in2:*
 - 36: gb_ba1:*
 - 37: gb_ba2:*
 - 38: em_hum3:*
 - 39: em_hum4:*
 - 40: gb_pr4:*
 - 41: gb_htg3:*
 - 42: gb_htg4:*
 - 43: gb_htg5:*
 - 44: gb_htg6:*

593
-1232
593

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	594	100.0	3402	5	AR047920	Sequence 1
2	590.8	99.5	3402	5	A48910	U24190 Trypanosoma
3	590.8	99.5	3402	35	TCU24190	AL022022 Mycobacte
4	45	7.6	47852	1	MTV023	L27277 Micrococcus
5	44	7.4	2402	1	MLORHO	AJ132828 Spermatoz
6	43.2	7.3	2028	7	SS1132828	M30933 E.tenella a
7	42.8	7.2	767	34	EIMMAX	D14486 Equine herp
8	41.6	7.0	5880	16	HSE4	U87960 Rattus norv
9	41.2	6.9	2479	12	RNU87960	AF020275 Homo sapi
10	40.4	6.8	477	11	AF020275	AL133030 Homo sapi
11	40.4	6.8	3932	10	HSR801297	X83546 R.norvegicu
12	40.4	6.8	3941	12	RNLAR2	AL009198 Mycobacte
13	40.4	6.8	69350	1	MTV004	AP000556 Homo sapi
14	40.4	6.8	149618	10	AP000556	AP000557 Homo sapi
15	40.4	6.8	150036	10	AP000557	AP000552 Homo sapi
16	40.4	6.8	157086	10	AP000552	AC007957 Homo sapi
17	40.4	6.8	158529	44	AC007957	AC009516 Homo sapi
18	40.4	6.8	169237	40	AC009516	AF139019 Cepaea ne
19	40	6.7	624	35	AF139019	X17207 Chlamydomon
20	39.6	6.6	618	7	CRGCR1	AF087653 Homo sapi
21	39.4	6.6	1442	40	AF087653	L38713 Gallus gall
22	39.2	6.6	818	4	CHKPROTAMI	M28100 Gallus gall
23	39.2	6.6	1528	4	CHKPROTAMN	X95503 M.musculus
24	39.2	6.6	2771	12	MMZINCFPR	X95504 M.musculus
25	39.2	6.6	3732	12	MMZINCFIP	AB018491 Homo sapi
26	39	6.6	259	9	AB018491	AB007820 Homo sapi
27	38.8	6.5	720	9	AB007820	X12735 Barley Cab-
28	38.8	6.5	1030	7	HVCAB2	X68361 M.fascicula
29	38.8	6.5	2858	10	MFAPOA4A	M15764 D.melanogas
30	38.4	6.5	150	34	DROPNFS1	M62760 Homo sapien
31	38.4	6.5	2268	9	HUMCOUPII	X61945 D.melanogas
32	38.4	6.5	2722	34	DMNG2	AB010711 Mus sp. m
33	38.4	6.5	2837	12	AB010711	AF030376 Acidamino
34	38.4	6.5	3015	2	AF030576	M23222 D.melanogas
35	38.4	6.5	5632	34	DROFSHB	M23221 D.melanogas
36	38.4	6.5	7792	34	DROFSHA	AC013031 Drosophil
37	38.4	6.5	11196	43	AC013031	AC014003 Drosophil
38	38.4	6.5	37518	43	AC014003	AL024484 Drosophil
39	38.4	6.5	39565	34	DMC96G10	I17281 Sequence 7
40	38.2	6.4	2943	5	I17281	M57388 Simian herp
41	38.2	6.4	2943	16	HSVGBQ	AF126748 Homo sapi
42	38	6.4	1159	40	AF126748	AF126749 Homo sapi
43	38	6.4	1472	40	AF126749	Y08701 M.musculus
44	38	6.4	2813	12	MMPININ	M33782 Human IFEB
45	37.8	6.4	1526	9	HUMTFEB	

ALIGNMENTS

RESULT 1
AR047920
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS

AR047920
Sequence 1
AR047920
AR047920.1

3402 bp
from patent US 5820864.
GI:5970263

DNA
PAT
29-SEP-1999

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 3402)
AUTHORS Paranhos-Baccala, G., Lesenechal, M. and Jolivet, M.
TITLE Trypanosoma cruzi antigen, gene encoding therefor and methods of
detecting and treating chagas disease
JOURNAL Patent: US 5820864-A 1 13-OCT-1998;
FEATURES Location/Qualifiers
source
1. 3402
BASE COUNT 888 a 821 c 956 g 737 t
ORIGIN
Query Match 100.0%; Score 594; DB 5; Length 3402;
Best Local Similarity 100.0%; Pred. No. 4.4e-117;
Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAGGTACAGCGTACGGCTTTTGGCTTCAATCGTACAGCCGACGGTAGCTGGCTCGCT 60
Db 1232 CAGGTACAGCGTACGGCTTTTGGCTTCAATCGTACAGCCGACGGTAGCTGGCTCGCT 1291
QY 61 GACATGTCGATTGACGATCTTCATCTCCGGTCTCCCGCAGGGAAGACACAG 120
Db 1292 GACATGTCGATTGACGATCTTCATCTCCGGTCTCCCGCAGGGAAGACACAG 1351
QY 121 CCAGGCCAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTCTCGGGCACT 180
Db 1352 CCAGGCCAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTCTCGGGCACT 1411
QY 181 GACGCGGAGTAGCAGTACATACCAATACGACTTCTCCGCTGTGTGATCCCTGCATCA 240
Db 1412 GACGCGGAGTAGCAGTACATACCAATACGACTTCTCCGCTGTGTGATCCCTGCATCA 1471
QY 241 CCCCTGTTTCAGCGCCAGCAAGCGCGCTCTTCCGCGGCGCATCGCTGAG 300
Db 1472 CCCCTGTTTCAGCGCCAGCAAGCGCGCTCTTCCGCGGCGCATCGCTGAG 1531
QY 301 CCGCAGCTGGGGACAGATCATTTGCTAAATCTAGTGAATCAGCTGGGGATTAAATGTCA 360
Db 1532 CCGCAGCTGGGGACAGATCATTTGCTAAATCTAGTGAATCAGCTGGGGATTAAATGTCA 1591
QY 361 CAAAGGAGCGTCTCAGCACTGGAGCGCCGCGCACAGGAGTCTACGGCGGTGACCTCC 420
Db 1592 CAAAGGAGCGTCTCAGCACTGGAGCGCCGCGCACAGGAGTCTACGGCGGTGACCTCC 1651
QY 421 ACAGTACCGCCCGCAGCGACCAAGTCCATACGGGCAATGGCCGACCTGTGACGGCT 480
Db 1652 ACAGTACCGCCCGCAGCGACCAAGTCCATACGGGCAATGGCCGACCTGTGACGGCT 1711
QY 481 GGATTGGTGAGTAATAGTGGTCCAGCGCGCTCTGCTCCACAGCGCGCGGCGAAA 540
Db 1712 GGATTGGTGAGTAATAGTGGTCCAGCGCGCTCTGCTCCACAGCGCGCGGCGAAA 1771
QY 541 CCAACAGGAGAGAAAGGCTCCCGGCGATGTGAACAGAGCTCCGTGGCGATA 594
Db 1772 CCAACAGGAGAGAAAGGCTCCCGGCGATGTGAACAGAGCTCCGTGGCGATA 1825
RESULT 2
LOCUS A48910 3402 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 1 from Patent WO9605312.
ACCESSION A48910
VERSION A48910.1 GI:2302570
KEYWORDS
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE 1 (bases 1 to 3402)
AUTHORS Paranhos-Baccala, G., Lesenechal, M. and Jolivet, M.

TITLE NOVEL TRYPANOSOMA CRUZI ANTIGEN, AND GENE CODING THEREFOR; THEIR
APPLICATION TO THE DETECTION OF CHAGAS' DISEASE
JOURNAL Patent: WO 9605312-A 1 22-FEB-1996;
COMMENT BIO MERIEUX (FR)
Other publication CA 2173957 960222
Other publication AU 3169195 960307
Other publication FR 2723589 960216.
FEATURES Location/Qualifiers
source
1. 3402
BASE COUNT 889 a 818 c 958 g 737 t
ORIGIN
Query Match 99.5%; Score 590.8; DB 5; Length 3402;
Best Local Similarity 99.7%; Pred. No. 2.1e-116;
Matches 592; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CAGGTACAGCGTAAACGGCTTTTGGCTTCAATCGTACAGCCGACGGTAGCTCGCTCGCT 60
Db 1232 CAGGTACAGCGTAAACGGCTTTTGGCTTCAATCGTACAGCCGACGGTAGCTCGCTCGCT 1291
QY 61 GACATGTCGATTGACGATCTTCATCTCCGGTCTCCCGCAGGGAAGACACAG 120
Db 1292 GACATGTCGATTGACGATCTTCATCTCCGGTCTCCCGCAGGGAAGACACAG 1351
QY 121 CCAGGCCAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTCTCGGGCACT 180
Db 1352 CCAGGCCAAAACATCGGTAGTGGCGACGGCGAAACCGGGGTGTGTCTCGGGCACT 1411
QY 181 GACGCGGAGTAGCAGTACATACCAATACGACTTCTCCGCTGTGTGATCCCTGCATCA 240
Db 1412 GACGCGGAGTAGCAGTACATACCAATACGACTTCTCCGCTGTGTGATCCCTGCATCA 1471
QY 241 CCCCTGTTTCAGCGCCAGCAAGCGCGCTCTTCCGCGGCGCATCGCTGAG 300
Db 1472 CCCCTGTTTCAGCGCCAGCAAGCGCGCTCTTCCGCGGCGCATCGCTGAG 1531
QY 301 CCGCAGCTGGGGACAGATCATTTGCTAAATCTAGTGAATCAGCTGGGGATTAAATGTCA 360
Db 1532 CCGCAGCTGGGGACAGATCATTTGCTAAATCTAGTGAATCAGCTGGGGATTAAATGTCA 1591
QY 361 CAAAGGAGCGTCTCAGCACTGGAGCGCCGCGCACAGGAGTCTACGGCGGTGACCTCC 420
Db 1592 CAAAGGAGCGTCTCAGCACTGGAGCGCCGCGCACAGGAGTCTACGGCGGTGACCTCC 1651
QY 421 ACAGTACCGCCCGCAGCGACCAAGTCCATACGGGCAATGGCCGACCTGTGACGGCT 480
Db 1652 ACAGTACCGCCCGCAGCGACCAAGTCCATACGGGCAATGGCCGACCTGTGACGGCT 1711
QY 481 GGATTGGTGAGTAATAGTGGTCCAGCGCGCTCTGCTCCACAGCGCGCGGCGAAA 540
Db 1712 GGATTGGTGAGTAATAGTGGTCCAGCGCGCTCTGCTCCACAGCGCGCGGCGAAA 1771
QY 541 CCAACAGGAGAGAAAGGCTCCCGGCGATGTGAACAGAGCTCCGTGGCGATA 594
Db 1772 CCAACAGGAGAGAAAGGCTCCCGGCGATGTGAACAGAGCTCCGTGGCGATA 1825
RESULT 3
LOCUS TCU24190
DEFINITION Trypanosoma cruzi 3402 bp mRNA INV 04-AUG-1997
ACCESSION U24190
VERSION U24190.1
KEYWORDS
SOURCE Trypanosoma cruzi.
ORGANISM Trypanosoma cruzi.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma; Schizotrypanum.

REFERENCE 1 (bases 1 to 3402)
AUTHORS Lesenechal, M., Duret, L., Cano, M.I., Mortara, R.A., Jolivet, M., Camargo, M.E., da Silveira, J.F. and Paranhos-Baccala, G.
TITLE Cloning and characterization of a gene encoding a novel immunodominant antigen of Trypanosoma cruzi
JOURNAL Mol. Biochem. Parasitol. 87 (2), 193-204 (1997)
MEDLINE 97391123
REFERENCE 2 (bases 1 to 3402)
AUTHORS Lesenechal, M., Franco Da Silveira, J., Mortara, R.A., Duret, L., Camargo, M.E., Jolivet, M. and Paranhos-Baccala, G.
TITLE Direct submission
JOURNAL Submitted (06-APR-1995) Mylene Lesenechal, BioMerieux, 22 rue Saint Jean de Dieu, Lyon 69007, France
FEATURES
source
1..3402
/organism="Trypanosoma cruzi"
/strain="G"
/db_xref="taxon:5693"
/dev_stage="epimastigote"
1..35
/gene="Tc40"
/note="spliced leader"
1..3402
/gene="Tc40"
266..3013
/gene="Tc40"
/codon_start=1
/product="Tc40 antigen"
/protein_id="AAC47657.1"
/db_xref="GI:790646"
/translation="MTVTVDLFNHPKSNRGRVMSVDAATFNEVPEAORVLADSOQV
LAYTKRHLVLRVYKRSNLLGKTPRAHAKPHAYKFNYSRNVASAGKGEFFVWVIT
DETERASGKPDAAKLTVKYFKLQDPVTPICFSEFFINAEORSORPLLVIETQAAILD
SSLRDETVESLEATLQNRCTTLTPQVSENLSVSGGWNFTTETPTWAACT
LNRSTPWAACGEFVKLLNLDTAVERNLSVLAASSTKGYIOWLLFGVAEPNLLRK
FVIGDSIVAMSSRETFVDFDRKALVNMHSPHNFCTHYMPCQVORNGFCFNRT
ADGCVLADMSNRLTFIHLRSRREOQOKTSVVATAKPCVCSGGDAASSHTNT
TSAAASPASPVSPAKAAPPAAARSAPAEHVGSKIIANLVNGLINVTQSRVYSTG
APATRSATVNTTAPORTSPYGNHGRPVTLGLVAANGASASPSPTAAKPTGEEK
ASAACSTVAINATPALHNASLPDPTGLVAAAAYOGEVHOSLERFESVITN
SRVLLKLPDTRRDEHQLNLEAQMTELOQSRPTQTPRDTSSAKSFVETTLV
LIADLSRNITKVRGVNEALHLDHEVRAIGNRLQTKNLIKSRDLAEALKEST
TOFTAQLQTVENLVKRELAEVLSINGSLTSLVENASLOKELINSIMSGVLDMMR
MREELCTLRSAKPKATMPDSSLHATSFQGRSAPETILATLALSMRYEQYRGLE
YMLAQCSLLRLFLSILITRENENAYSELIVETPNVDVWCVSLLQLLEAAATEAEKE
VVGVVAIDILSDQIAQNGALGSKLTMTAMRAFERQARSETTSRFLQCLNLEKLLO
S"
BASE COUNT 889 a 818 c 958 g 737 t
ORIGIN

Query Match 99.5%; Score 590.8; DB 35; Length 3402;
Best Local Similarity 99.7%; Pred. No. 2.1e-116;
Matches 592; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGTACAGGTAAACGGCTTTTCCTTCATCTGACAGCGACGGTAGCTGCGTCTGGCT 60
|||||
Db 1232 CAGGTACAGGTAAACGGCTTTTCCTTCATCTGACAGCGACGGTAGCTGCGTCTGGCT 1291
QY 61 GACATGTCGATTCGATTGACGATCTCCACTCTCCGCTCCCGCGAGGAAGAACAGCAG 120
|||||
Db 1292 GACATGTCGATTCGATTGACGATCTCCACTCTCCGCTCCCGCGAGGAGAACAGCAG 1351
QY 121 CCAGGCCAAAAACATCGGTAGTGGGACGGCGAAACCGGGGTGTGTGCTCTCGGCACCT 180
|||||
Db 1352 CCAGGCCAAAAACATCGGTAGTGGGACGGCGAAACCGGGGTGTGTGCTCTCGGCACCT 1411
QY 181 GAGCGCGGAGTAGCAGTATACCAATACGACTTTCGCCGCTGCTGATCCCCCTGCATCA 240
|||||
Db 1412 GAGCGCGGAGTAGCAGTATACCAATACGACTTTCGCCGCTGCTGATCCCCCTGCATCA 1471
QY 241 CCCCTCTGTTTACGCGCAGGACGCGCGCTCTCTCGCGCGGCGATCGGCTGAG 300
|||||

Db 1472 CCCCTGTTTTCAGCGCCAGCGAAGCAGCGCGCTCTCTCGCGCGCGGATCGGCTGAG 1531
QY 301 CGGACAGTGGGAGCAAGATCATTTCTATCTAGTGAATCATCTGGGATTAATGTACCC 360
|||||
Db 1532 CGGACAGTGGGAGCAAGATCATTTCTATCTAGTGAATCATCTGGGATTAATGTACCC 1591
QY 361 CAAAGGAGCGTCTGACGACTGGAGCGCGGCGCACGACGAGTCTACGGCGGTGACGCTCC 420
|||||
Db 1592 CAAAGGAGCGTCTGACGACTGGAGCGCGGCGCACGACGAGTCTACGGCGGTGACGCTCC 1651
QY 421 AGCACTACCGCCCGCCAGCGAACAAGTCCATACGGGCAACATGGCGGACCTGTGACGGCT 480
|||||
Db 1652 AGCACTACCGCCCGCCAGCGAACAAGTCCATACGGGCAACATGGCGGACCTGTGACGGCT 1711
QY 481 GGATTGGTGGCAGCTAATAGTGTGCGGCGGCGCTCTCTCCACACCGCGGGGAAA 540
|||||
Db 1712 GGATTGGTGGCAGCTAATAGTGTGCGGCGGCGCTCTCTCTCCACACCGCGGGGAAA 1771
QY 541 CCAACAGGAGAGAAAGAGCCCTCCGCGCATGTGAACGAGCTCGTGGCGATA 594
|||||
Db 1772 CCAACAGGAGAGAAAGAGCCCTCCGCGCATGTGAACGAGCTCGTGGCGATA 1825

RESULT 4
MTV023 47852 bp DNA BCT 17-JUN-1998
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 148/162.
DEFINITION AL022022 AL123456
ACCESSION AL022022.1 GI:3261554
VERSION
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Mycobacterium tuberculosis.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium.
REFERENCE 1 (bases 1 to 47852)
AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, K., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G.
TITLE Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
JOURNAL Nature 393 (6685), 537-544 (1998)
MEDLINE 98295987
REMARK Erratum: [[published erratum appears in Nature 1998 Nov 12:396(6707):190]]
REFERENCE 2 (bases 1 to 47852)
AUTHORS Parkhill, J.
JOURNAL Direct Submission
TITLE Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhillesanger.ac.uk
COMMENT On Jun 27, 1998 this sequence version replaced gi:2924430.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBPase (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome

binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

```
FEATURES
    source          Location/Qualifiers
    1..47852
        /organism="Mycobacterium tuberculosis"
        /strain="H37Rv"
        /db_xref="taxon:1773"
    <1..18
        /organism="Mycobacterium tuberculosis"
        /strain="H37Rv"
        /db_xref="taxon:1773"
        /clone="Y13E12"
        complement(1..1695)
        /gene="Rv3494c"
        /note="Rv3494c, (MTV023.01c), len: 564. Unknown Pro-rich protein similar to several Mycobacterium tuberculosis proteins e.g. MTCI28.14 (515 aa), MTCY19H5.28c (516 aa) and (MTV051.09). Has hydrophobic stretch, possibly signal peptide at N-terminus. FASTA scores: 297050|MTCI28_14 (515 aa) opt: 1006 z-score: 582.8 E(): 5.9e-25; 34.3% identity in 559 aa overlap: 297182|MTCY19H5_28 (516 aa) opt: 979 z-score: 567.7 E(): 4.1e-24; 33.5% identity in 555 aa overlap. TBparse score is 0.897"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein Rv3494c"
        /protein_id="CAAL7731.1"
        /db_xref="GI:2924431"
        /db_xref="SPTREMBL:O53539"
        /translation="MDRLAKIQLSIFAVITVTLVSMAIFYLRLPATRGIGTYGVA DFVAGGLYKNANTYRGVAGVSGVGLNPGVTAHRLNSGTATPSNVTATRSVS AIGEQLIDLPNPNPSTKLNGFRIGRQNRIGQVADLLROAELLGSLGTRURE LSHLEATNAGPEARLIESARLLVDANANYPOVSOLIDQAGPFLQAIKAGGDI KSLADLRTWLRAADRLDRLADAPDAIDEANTAFSGIRSPFALAASLANLGR VGIYHKSIEQLVFPALFAALITISAGGVQDEGAKLDFKIDLHDPPECMGTFLPFP LVRSDEIREIPDMYCKTAQNDPSTVGRARNYPCOEPGKRAPTVOLCRDPRGYV PVTGNWRGPIPYGTEVTDGRLNLPNKEFPIPGCADDPGVPIGVPPPPQVAGPG PAPHQAPAPNDPNPPPPFSSWPPGPIPEPPIATIPPPPPPEGGPPPGP APQPQASGPATIIIDQLSGAFADPAGGTGIFAPGTMGASSAENWVLDLRDPRQL"
        complement(1..1695)
        /gene="Rv3494c"
    <1..47852
        /note="fragment designated v023. Does not represent a physical clone"
        complement(1703..1707)
        /note="possible RBS for Rv3494c"
        complement(1706..2860)
        /gene="lprN"
        complement(1706..2860)
        /gene="lprN"
        /note="Rv3495c, (MTV023.02c), len: 384. lprN, similar to Mycobacterium tuberculosis proteins MTCI28.13 (390 aa) and MTCY19H5.29 (402 aa) and (MTV051.08). Probably lipo-protein, contains possible signal sequence and appropriately positioned PS00013 Prokaryotic membrane lipo-protein lipid attachment site. FASTA scores: 91|297050|MTCI28_13 (390 aa) opt: 653 z-score: 762.1 E(): 0; 33.6% identity in 363 aa overlap; and 297182|MTCY19H5_29 (402 aa) opt: 572 z-score: 667.9 E(): 1.1e-29; 31.8% identity in 362 aa overlap. TBparse score is 0.897"
        /codon_start=1
        /transl_table=11
        /product="lprN"
        /protein_id="CAAL7732.1"
        /db_xref="GI:2924432"
        /db_xref="SPTREMBL:O53540"
        /translation="MNRILWRAIILTASSALLAGCOFGNLPLPCTAGHGEYASV TVEMADYALPQNSPVLWPDVTVGSVAGIVAVORPGSYFAVKLLDRLDNLDPANAV AKVQSYLGLSHLVEAPDTPPTGRLVDSRTEANTDRPTTEVFSALGVVYWK GNVALEETIDETHVAGRQKQFVNLVPLRLAELTAGLNROVHDIIDLADGNRVSAL LARDKDNLRALDTLPDAVRVLRNQNRDHIVDFAALKRLTMTVSHVLAETKYDFGDEL
```

```

KDLYSIVKALNDRDKDFTSLQLLTFPPFPNGFIKQAVRGDYLNVFTTDLTLRRIGE
TFTTAYFDPNNAHDEILNPDFFLIGELANLSGQADDFKIPPGTASQ"
complement(2798..2830)
/gene="lprN"
/note="PS00013 Prokaryotic membrane lipo-protein lipid
attachment site"
complement(2857..4212)
/gene="Rv3496c"
complement(2857..4212)
/gene="Rv3496c"
/note="Rv3496c, (MTV023.03c), len: 451. Unknown but
similar to Mycobacterium tuberculosis proteins MTCI28.12
(530 aa), MTCY19H5.30c (508 aa) (MTV051.07). Hydrophobic
region at N-terminus. FASTA scores: 297050|MTCI28_12 (530
aa) opt: 838 z-score: 711.0 E(): 4.3e-32; 33.1% identity
in 473 aa overlap; and 297182|MTCY19H5_30 (508 aa) opt:
821 z-score: 697.0 E(): 2.6e-31; 35.1% identity in 453 aa
overlap. TBparse score is 0.891"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3496c"
/protein_id="CAAL7733.1"
/db_xref="GI:2924433"
/db_xref="SPTREMBL:O53541"
/translation="MMGRVAMLTGSRGLRYATVIALVAALVGVVYVLSGTGKRTIVG
YFTSAVLPGQQRVGLVPGVEIDMIERSDSVKITMSVDKVPDQVIMSPN
LVAARFIQLTPTVGTGAVLPDNGRIDLRTAVPVEWDEKGLTAAALSPAAGELQ
GLGAALINQADTLIDNGDSLHNLAEQAQVAGRLGDSRGDIFGTVKNQVLDALSE
SDEQIVQFAGHAVSVQSLVADSANLDOTLGTLNQALSDIRGLFNNSNLTLETNQL
NDFAOITLDSQSENIQVHLVAGGPTNFYNIYDPAQGLNLGLSLINPFPANPVQFICGG
SFTDAAGSPADPYRRAETICRERLQVRLBTVNVPIMFHLNITATYKGOIIVDTP
ATEASETPVELTWTVPAGGAPVGNADQLSLVPPAPCAPAPAPGAPGGEHGGG
G"
complement(2867..2871)
/gene="Rv3496c"
/note="possible RBS for Rv3495c"
complement(4209..5282)
/gene="Rv3497c"
complement(4209..5282)
/gene="Rv3497c"
/note="Rv3497c, (MTV023.04c), len: 357. Unknown but
similar to Mycobacterium tuberculosis proteins MTCY19H5.31
(481 aa), MTCI28.11 (515 aa) and MTV051.06. Hydrophobic
region at N-terminus. FASTA scores: 297182|MTCY19H5_31
(481 aa) opt: 611 z-score: 667.7 E(): 1.1e-29; 32.5%
identity in 332aa overlap; and 297050|MTCI28_11 (515 aa)
opt: 587 z-score: 641.2 E(): 3.3e-28; 30.1% identity in
335 aa overlap. TBparse score is 0.889"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3497c"
/protein_id="CAAL7734.1"
/db_xref="GI:2924434"
/db_xref="SPTREMBL:O53542"
/translation="MLNRPKSSKHERDPLRTGIFGLVLVICVLIATAGSYGLFWPQGG
KYIDAFDTAGITFGNSVYVSLKVGAVSAVSLAGNSAKVIFSDRSIVGDQSLAA
IKTDILGRSIAVSPAGSGKSTTIPLSRTTPTTLNGVLDLRNANDLNRPQFQA
LNVFTQALHDATPQVRGADGLTSLRNLNRDEALQGLLAHAKSVTSYLSRAEQVN
KIVEDGNOLFALDARRAALSALISGIDDDVAAQISGFVADNRKEFGPALSKLNLVLN
LNERDYITETALKRLPTTATTLGVEVVGSGPFGFNVYSLVPLGVLATVDFLVQPKGL
PDLADYLRGFIQERWIRPKSP"
complement(5272..5324)
/gene="Rv3498c"
complement(5272..5324)
/gene="Rv3498c"
/note="Rv3498c, (MTV023.05c), len: 350. Unknown but
similar to Mycobacterium tuberculosis proteins MTCI28.10
(346 aa), MTCY19H5.32c (346 aa) and MTV051.05. Hydrophobic
region at N-terminus. FASTA scores: 297050|MTCI28_10 (346
aa) opt: 755 z-score: 883.6 E(): 0; 36.8% identity in
340 aa overlap; and 297182|MTCY19H5_32 Mycobacterium
tuberculosis cosmi (275 aa) opt: 699z-score: 820.1 E(): 0;
47.0% identity in 249 aa overlap. TBparse score is 0.878"
```


[illegible]

kinase gene
J. Gen. Virol. 71 (Pt 8), 1801-1805 (1990)
90362067
REFERENCE
3 (bases 1 to 5880)
Nicolson, L.
Unpublished (1993)
Submitted (10-SEP-1990) to DDBJ by:
Lesley Nicolson
Dept. Veterinary Pathology, University of Glasgow
vet School
Bearsden Roda, Glasgow G61 1QH, Scotland
UK.
Phone: 041-339-8855
Fax: 041-330-5733.

FEATURES

source
Location/Qualifiers
1..5880
/organism="Equine herpesvirus 4"
/db_xref="taxon:10331"
complement(99..104)
/note="put. polyadenylation signal (UL24); putative"
complement(115..120)
/note="put. polyadenylation signal (UL24); putative"
complement(116..134)
/note="putative; ORF1 (UL24 homologue)"
/codon_start=1
/protein_id="BAA03377.1"
/db_xref="GI:221821"
/translation="MKRKORLTARSLRAGIRCHNRFYNAVQDLSAKKNGVYGARL
APFLSELPATKLTAMGVSLAFYNLGRPDCTVQFGHSDAKGVCLIELKTC
RFSKNMTASLORLKHGMDLHSCRLIAKTLPPSGEILILAPLVFAVQGMVLR
VTRLSPQVYNAVLSTISRLAEPSPISERSRRRCVTRRTNSKAFRAKTTGSI
QPIIOAKPAATAVASLFSATAQAANTNNAVGYQATISLANPLAWASLAPK"
861..865
/note="put. TATA box (TK); putative"
887
/note="pot. transcription initiation site; putative"
949..2007
/partial
/codon="ORF2; (TK)"
/codon_start=1
/protein_id="BAA03378.1"
/db_xref="GI:221822"
/translation="NAACVPCEAPRSAGTPTREQTVIRIYLDGVYICKSTGRV
MASRAGSGPTLYFPPEWAYRTLETDVIGSIYDQNRKQCNLANVDAALITAHYQ
TFTPTLILHDTCTLFNGNSLQRTQDITLFDHPHVASTVCFPAARVLLGDMQ
CALMAVATLPREPQGNIVVITLNEEHRILRTRAGIQDITLIALRNVTFML
VNTCHFSLGRVYRWGDELPTSCGAYKHRAQMDAFQERVSPELGDTLFLFKTQEL
LDRGVILEVHAWALDALMLKRLNLNLFVSADLSGTTPRQCAAVVESLLPLMSLTSLDFD
SASALERAARTNAEMGV"
2049..2054
/note="put. polyadenylation signal (TK); putative"
2127..2133
/note="putative"
2197..2202
/note="putative"
2225..4792
/partial
/note="ORF3; (GH)"
/codon_start=1
/protein_id="BAA03379.1"
/db_xref="GI:221823"
/translation="MSQYLKAILVAATIVSAIPWVTPVSTSPQOTKLHYVNGT
WHNNTFNTRYDIMEPVNNSSTFFVAINSEFRNTVNTPLGASVFWLSKAL
NPPKQPCIANPEPDRGCNVSTSLFFNDNLEPFLMTKNLFEFLVDPNITGW
TPEKSTVATKGNPVGLSPPTSPDVNNTIRDDGTPKOHLSIIDHTFVLDLQNE
TKTLVISPFAWPTITAFHAGITVNGDITQAIJVLNGFMGLQISSVNPPLMIY
ANDVRAIRVNLPRRRLPEPGPAGPIYKYVYLSGDNFYLGHGMSKISRVAAYPE
ESLDRYHLSDNLTLAMLELSSGKSDSYIYRIIRIARLAVATFSLAEVIRLSDY
MLQFAIDINLRLITPLVMKYAAGTADSSYISDVAQDFEVAQKIEKIVADIN
IENELKRHLKSLKYVYSRKPLPNVAFNRLITAMYKEAKRQRTNWTMRE
VLFVAGAAAGSHVITDGPDLGLHAKDSSMFLSLNRLNLLCTAMCTASHAVAGV
KLEEVNAGIAGGVFSLLEVPSCMASRFDLAEHEHVDLLSLVPRLYDNLATGL
EDDGTTHSYGSRANGLNSRIAYNFDAVRVTPPELASCSTKLPKLVVPLASNRSI

polyA_signal

polyA_signal

CDS

Query Match 7.0%; Score 41.6; DB 16; Length 5880;
Best Local Similarity 44.1%; Pred. No. 12;
Matches 173; Conservative 0; Mismatches 219; Indels 0; Gaps 0;
QY 145 GCGACGGCGAAACCGGGGTGCTGCTCGGCACCTGACGCGCGAGTAGCAGTACATACC 204
DB 4893 GCGCGCGCGAGGGCTGCTCGCGCGGAGGGCTGCTCGCGCGGAGGGCTGCGCGCGGCG 4952
QY 205 AATAGACTTCGCGCTCTCATCCCTGCATCATACCCCTGTTTTCAGCGCCAGCAAG 264
DB 4953 GCGAGGGCTGCGCGCGGCGGAGGGCTGCGCGCGGCGGCGGAGGGCTGCGCGCGGCG 5012
QY 265 GCAGCGCGCGCTCTCGCGCGCGGCGATCGCTGAGCGCGCAGCTGGGGAGCAAGATCAT 324
DB 5013 GCGCGGAGGGCTGCGCGCGGCGGAGGGCTGCTCGCGCGGCGGAGGGCTGCTGCGG 5072
QY 325 GCTAATCTAGTGAATCAGCTGGGGATTAAATGTACACCAAGAGAGCTCTCAGCATGGA 384
DB 5073 GCGCGCGGAGGGCTGCTCGCGCGGCGGAGGGCTGCTGCGCGCGGCGGAGGGCTGCT 5132
QY 385 GCGCGCGCGCACACGAGGTCTACGCGGCTGACCTCACACCTACCGCGCGGAGCAACA 444
DB 5133 GCGCGCGCGGAGGGCTGCTCGCGCGGCGGAGGGCTGCTGCGCGCGGCGGAGGGCT 5192
QY 445 AGTCATACGGCACAAATGGCGGACCTGTGACGGCTGGATTGGTGGCAGCTAATAGTGT 504
DB 5193 GCTCGCGCGCGCGGAGGGCTGCTGCTGCGCGCGGCGGAGGGCTGCTGCTGCGCGGCG 5252
QY 505 GCGAGCGCGCTGCTGCTCCACAGCGCGGCG 536
DB 5253 GCGAGGGCTGCTGCGCGCGGCGGAGGGCTGCG 5284

RESULT 9

RN087960/2479 bp DNA ROD 04-MAR-1997
LOCUS Rattus norvegicus leukocyte common antigen receptor (LAR) gene,
trans-spliced alternative untranslated exon.
DEFINITION U87960
ACCESSION U87960
VERSION U87960.1 GI:1857717
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 2479)
AUTHORS Zhang, C., Zhang, J.S., Martignetti, J.A., Massa, S.M. and Longo, F.M.
TITLE Evidence for a mammalian trans-spliceon
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2479)
AUTHORS Zhang, C., Zhang, J.S., Martignetti, J.A., Massa, S.M. and Longo, F.M.
TITLE Direct Submission


```

/note="Rv3348, (MTV004.04), len: 163 aa; unknown,
identical to M. tuberculosis hypothetical protein
TR:P96234 (EMBL:783864) MTCY0A6.25 (163 aa). Partially
similar to several insertion elements, e.g.
P198341Y111_STRCL_INSERTION_ELEMENTIS116, (399 aa, 28.5%
identity in 158 aa overlap)."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3348"
/protein_id="CAAL5733.1"
/db_xref="GI:2661627"
/db_xref="SPTREMBL:P96234"
/transl_table="MTAENPGRSRTTLVGDAAITACHIAIRDVGARSIRFSVEPT
LAGRLTDLKSLGVDIDATVEPTSMTLPLTIAVENAGDTMHGARGHARGALV
GKSSDVIDAEVLTRASEVFDLTPLTPTLAQLALRSVIRAGAVIDANRSWRLMS
LAR"
complement(11239..11979)
/gene="Rv3349c"
CDS
complement(11239..11979)
/gene="Rv3349c"
/note="Rv3349c, (MTV004.05c), len: 246 probable
transposase pseudogene fragment, similar to part of
TR:O50911 (EMBL:U10634) IS204 PUTATIVE TRANSPOSASE from
NOCARDIA ASTEROIDES (377 aa), fasta scores; opt: 288
z-score: 322.6 E(): 8.3e-11, 48.5% identity in 97 aa
overlap"
/codon_start=1

Query Match      6.8%; Score 40.4; DB 1; Length 69350;
Best Local Similarity 53.1%; Pred. NO. 18;
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 142 GTGGCGACGGGCAACCGGGTGTCTCTGGGCATGTAGCGCGGAGTAGCTAGTAT 201
Db 37024 GCGGCGACGGGTGACGGGGGACGCGGACCGCGGTGACGGGTACTGGCGG 37083

QY 202 ACCAATACGACTCTCGCGCTGCTCATCCCTGCATCACCCCTGTTTCAGCGCCAGCC 261
Db 37084 GCACACGCGCAACTTCGGCAATGCGGTGGCGCGGTGACCGCGGCGGGGACCCC 37143

QY 262 AAGCGACGGCCCTCTCTCGCGCGCGGATGCGGTGAGCGCC 303
Db 37144 ATGGCACCCCGGACGCGCAACGCGGACCGCGGACCGCGCACCG 37185

RESULT 14
AP000556/c
LOCUS
DEFINITION Homo sapiens genomic DNA, chromosome 22q11.2, BCRL2 region,
clone:KB1172D5.
ACCESSION AP000556
VERSION AP000556.2 GI:6139049
KEYWORDS
SOURCE Homo sapiens DNA, clone:KB1172D5.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 149618)
Shimizu, N.
Human DNA sequence from clone KB1172D5 on chromosome 22q11.2
Published Only in Database (1999) In press
2 (bases 1 to 149618)
Shimizu, N.
Direct Submission
Submitted (01-OCT-1999) to the DDBJ/EMBL/GenBank databases.
Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-0015, Japan
(E-mail: nshimizu@db.med.keio.ac.jp, Tel: 81-3-3351-2370,
Fax: 81-3-3351-2370)
On Oct 29, 1999 this sequence version replaced gi:6006352.
This is a complete sequence of the insert of KB1172D5 clone. The
proximal adjacent clone is KB1183D5 (Acc.#AP000552) with 67422-bp
overlapping. The distal adjacent clone is KB1323B2 (Acc.#AP000557)

```

```

with 69793-bp overlapping
Sequence updated (26-Oct-1999).
Location/Qualifiers
source 1. 149618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="KB1172D5"
/map="22q11.2"
complement(1..667)
/rpt_family="L1M4c"
repeat_region
evidence-not_experimental
complement(1010..1716)
/rpt_family="L1M4c"
repeat_region
evidence-not_experimental
complement(1717..2021)
/rpt_family="AluSx"
repeat_region
evidence-not_experimental
complement(2022..2297)
/rpt_family="L1M4c"
repeat_region
evidence-not_experimental
2297..2368
/rpt_family="L1MB4"
repeat_region
evidence-not_experimental
2760..3043
/rpt_family="AluSq"
repeat_region
evidence-not_experimental
3044..3066
/rpt_family="(TAAA)n"
repeat_region
evidence-not_experimental
complement(4122..4305)
/rpt_family="AluSg/x"
repeat_region
evidence-not_experimental
4949..5151
/rpt_family="AluJb"
repeat_region
evidence-not_experimental
5649..5748
/rpt_family="MIR"
repeat_region
evidence-not_experimental
complement(8880..9173)
/rpt_family="AluY"
repeat_region
evidence-not_experimental
10813..10860
/rpt_family="GC-rich"
repeat_region
evidence-not_experimental
10862..10892
/rpt_family="(CGG)n"
repeat_region
evidence-not_experimental
11091..11118
/rpt_family="GC-rich"
repeat_region
evidence-not_experimental
complement(11968..12299)
/rpt_family="AluJo"
repeat_region
evidence-not_experimental
13437..13529
/rpt_family="MIR"
repeat_region
evidence-not_experimental
15901..16312
/rpt_family="MER31B"
repeat_region
evidence-not_experimental
16308..16476
/rpt_family="MER31B"
repeat_region
evidence-not_experimental
complement(16477..16655)
/rpt_family="AluSc"
repeat_region
evidence-not_experimental
17091..17265
/rpt_family="L2"
repeat_region
evidence-not_experimental
17219..17272
/rpt_family="MIR"
repeat_region
evidence-not_experimental
17597..17671

```

```
/rpt_family="G-rich"
/evidence-not_experimental
complement(17672..17831)
/rpt_family="AluSx"
/evidence-not_experimental
17832..17880
/rpt_family="MER45C"
/evidence-not_experimental
21993..22302
/rpt_family="AluY"
/evidence-not_experimental
22304..22737
/rpt_family="L1MD3"
/evidence-not_experimental
22888..23013
/rpt_family="AluYb8"
/evidence-not_experimental
23014..23323
/rpt_family="AluY"
/evidence-not_experimental
23324..23496
/rpt_family="AluYb8"
/evidence-not_experimental
complement(24098..24406)
/rpt_family="AluYb"
/evidence-not_experimental
complement(24447..24558)
/rpt_family="MIR"
/evidence-not_experimental
24631..24721
/rpt_family="CT-rich"
/evidence-not_experimental
24952..25245
/rpt_family="AluSx"
/evidence-not_experimental
26011..26140
/rpt_family="L2"
/evidence-not_experimental
26310..26339
/rpt_family="(TTTG)n"
/evidence-not_experimental
complement(26340..26623)
/rpt_family="AluSx"
/evidence-not_experimental
26731..26833
/rpt_family="L2"
/evidence-not_experimental
complement(26798..26898)
/rpt_family="L2"
/evidence-not_experimental
26899..27029
/rpt_family="FLAM_C"
/evidence-not_experimental
complement(27037..27337)
/rpt_family="AluY"
/evidence-not_experimental
complement(27338..27380)
/rpt_family="L2"
/evidence-not_experimental
27868..28160
/rpt_family="AluY"
/evidence-not_experimental
28255..28286
/rpt_family="(TTTTA)n"
/evidence-not_experimental
complement(28298..28602)
/rpt_family="AluYb"
/evidence-not_experimental
complement(28612..28912)
/rpt_family="AluSx"
/evidence-not_experimental
complement(28927..29192)
/rpt_family="AluSg"
```

```
/evidence-not_experimental
complement(29229..29542)
/rpt_family="AluSx"
/evidence-not_experimental
complement(29544..29741)
/rpt_family="7SLRNA"
/evidence-not_experimental
29808..30107
/rpt_family="AluSx"
/evidence-not_experimental
complement(30206..30372)
/rpt_family="L2"
/evidence-not_experimental
complement(30521..30840)
/rpt_family="AluY"
/evidence-not_experimental
complement(30841..30969)
/rpt_family="AluYb"
/evidence-not_experimental
complement(31012..31228)
/rpt_family="AluJo"
/evidence-not_experimental
31307..31387
/rpt_family="MIR"
/evidence-not_experimental
31540..31819
/rpt_family="AluSc"
/evidence-not_experimental
```

Query Match 6.8%; Score 40.4; DB 10; Length 149618;
Best Local Similarity 52.9%; Pred. No. 17;
Matches 109; Conservative 0; Mismatches 96; Indels 1; Gaps 1;

```
QY 141 AGTGGCAGCGGAAACCGGGGTGTGTCTCGGGCAGTCAGCGCGGAGTAGCAGTCA 200
|| ||| ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35731 AGCAGCGCGGCTGCAGCTGTGTCTCGGCTCTCTCCACAGTCGCGCGCGCGCGCC 35672

QY 201 TACCAATAGACTTTCGCGCTGCTGCATCCCTGCATCACCCCTGTTTCAGCGCAGC 260
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35671 TCGTGCCCGAGCTCTCGGCATCGCGCGCGCTGCTCGCGCTCCAGCGCCAGC 35612

QY 261 CAAGGAGCGCGCGCTCTCGCGCGCGGATC-GGCTGAGCGCGCAGCTGGGAGCAAGA 319
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35611 ATCTGCGCTCGCACCTCTCGGAGTTGGCTCTGCGCGCGCGCTCGGCGCCAGCGCCA 35552

QY 320 TCATTGCTAATCTAGTGAATCAGCTG 345
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35551 CGTTGCTGCTGAAGCATCAACTG 35526
```

RESULT 15

AP000557
LOCUS AP000557 150036 bp DNA PRI 02-NOV-1999
DEFINITION Homo sapiens genomic DNA, chromosome 22q11.2, BCRL2 region,
clone:KB1323B2.

ACCESSION AP000557

VERSION AP000557.2 GI:6174873

KEYWORDS

SOURCE Homo sapiens DNA, clone:KB1323B2.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

Shimizu,N.

REFERENCE 1 (bases 1 to 150036)

KB1323B2

Homo sapiens genomic DNA, chromosome 22q11.2, BCRL2 region, clone

Published Only in DataBase (1999) In press

Shimizu,N.

REFERENCE 2 (bases 1 to 150036)

Shimizu,N.

Direct Submission

Submitted (01-OCT-1999) to the DDBJ/EMBL/GenBank databases.

Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular

Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan

(E-mail:nshimizudmb.med.keio.ac.jp, Tel:81-3-3351-2370;
Fax:81-3-3351-2370)
On Nov 2, 1999 this sequence version replaced gi:6006353.
This is a complete sequence of the insert of KB1323B2 clone. The
proximal adjacent clone is KB1172D5 (Acc.#AP000556) with 69794-bp
overlapping. The distal adjacent clone is KB1802C5 (Acc.#AP000558)
with 21623-bp overlapping.
Sequence updated (29-Oct-1999).

COMMENT

FEATURES

Source	Location/Qualifiers
repeat_region	1..150036 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22" /clone="KB1323B2" /map="22q11.2" 519..594
repeat_region	/rpt_family="(TTTG)n" /evidence=not_experimental complement(554..565)
repeat_region	/rpt_family="FLAM_C" /evidence=not_experimental 566..594
repeat_region	/rpt_family="(TTTG)n" /evidence=not_experimental complement(595..716)
repeat_region	/rpt_family="FLAM_C" /evidence=not_experimental 813..1106
repeat_region	/rpt_family="AluSx" /evidence=not_experimental 1918..2204
repeat_region	/rpt_family="AluJb" /evidence=not_experimental 2613..2915
repeat_region	/rpt_family="AluSp" /evidence=not_experimental complement(3253..3353)
repeat_region	/rpt_family="L1ME3A" /evidence=not_experimental complement(3396..3501)
repeat_region	/rpt_family="L1ME3A" /evidence=not_experimental complement(3779..4072)
repeat_region	/rpt_family="AluSx" /evidence=not_experimental 4132..4429
repeat_region	/rpt_family="AluSx" /evidence=not_experimental complement(5553..5635)
repeat_region	/rpt_family="MER5A" /evidence=not_experimental complement(5930..6220)
repeat_region	/rpt_family="AluY" /evidence=not_experimental 7330..7497
repeat_region	/rpt_family="MIR" /evidence=not_experimental complement(9725..9855)
repeat_region	/rpt_family="L1ME3A" /evidence=not_experimental 9901..10191
repeat_region	/rpt_family="AluYb8" /evidence=not_experimental 17988..18048
repeat_region	/rpt_family="(TC)n" /evidence=not_experimental complement(22321..22478)
repeat_region	/rpt_family="HAL1" /evidence=not_experimental 22670..22733
repeat_region	/rpt_family="(TA)n" /evidence=not_experimental complement(22743..22997)
repeat_region	/rpt_family="AluSx" /evidence=not_experimental complement(22998..23084)
repeat_region	/rpt_family="MER58A" /evidence=not_experimental 23091..23406
repeat_region	/rpt_family="AluSg" /evidence=not_experimental 23407..23550
repeat_region	/rpt_family="AluSg" /evidence=not_experimental 23612..23757
repeat_region	/rpt_family="AluSg" /evidence=not_experimental 23758..24046
repeat_region	/rpt_family="AluY" /evidence=not_experimental 24047..24058
repeat_region	/rpt_family="AluSg" /evidence=not_experimental 24059..24357
repeat_region	/rpt_family="AluSx" /evidence=not_experimental 24358..24506
repeat_region	/rpt_family="AluSg" /evidence=not_experimental complement(24507..24815)
repeat_region	/rpt_family="AluY" /evidence=not_experimental 24816..24846
repeat_region	/rpt_family="AluSg" /evidence=not_experimental complement(25153..25450)
repeat_region	/rpt_family="AluJb" /evidence=not_experimental 26278..26587
repeat_region	/rpt_family="AluSx" /evidence=not_experimental complement(26842..26972)
repeat_region	/rpt_family="L2" /evidence=not_experimental 26985..27136
repeat_region	/rpt_family="AluJb" /evidence=not_experimental complement(27163..27322)
repeat_region	/rpt_family="L2" /evidence=not_experimental complement(27951..28039)
repeat_region	/evidence=not_experimental complement(28041..28351)
repeat_region	/rpt_family="AluY" /evidence=not_experimental 28352..28383
repeat_region	/rpt_family="(TTTG)n" /evidence=not_experimental complement(28384..28664)
repeat_region	/rpt_family="AluJb" /evidence=not_experimental complement(28669..28964)
repeat_region	/rpt_family="AluJo" /evidence=not_experimental complement(28968..29280)
repeat_region	/rpt_family="AluJb" /evidence=not_experimental 29359..29484
repeat_region	/rpt_family="AluSg" /evidence=not_experimental complement(29487..29566)
repeat_region	/rpt_family="AluY" /evidence=not_experimental 29606..29747
repeat_region	/rpt_family="FLAM_C"

```
/evidence-not_experimental
29750. .30039
/rpt_family="AluSq"
/evidence-not_experimental
30042. .30356
/rpt_family="AluJo"
/evidence-not_experimental
complement(30945. .31067)
/rpt_family="MLTIC"
/evidence-not_experimental
complement(31068. .31371)
/rpt_family="AluYb8"
/evidence-not_experimental
complement(31372. .31489)
/rpt_family="MLTIC"
/evidence-not_experimental
complement(31508. .31792)
/rpt_family="AluJb"
/evidence-not_experimental
31992. .32012
/rpt_family="Alu_rich"
/evidence-not_experimental
complement(32013. .32138)
/rpt_family="FLAM_C"
/evidence-not_experimental
32185. .32279
/rpt_family="L1M4"
/evidence-not_experimental
32828. .32967
/rpt_family="AluJb"
/evidence-not_experimental
complement(32986. .33124)
/rpt_family="MLTIIH"

Query Match          6.88;  Score 40.4;  DB 10;  Length 150036;
Best Local Similarity 52.98;  Pred. No. 17;
Matches 109;  Conservative 0;  Mismatches 96;  Indels 1;  Gaps 1;

QY  141 AGTGGCGACGGCGGAACGGGGGTGTGTCTCGGGCACTGACGGCGGAGTAGCAGTCA 200
    || ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119367 AGCAGCGCGCGCTGCAGCTGTGTCTCGGCTCTCTCGCCACGTCGCGCGCGCGCC 119426
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  201 TACCAATACGACTTCTGCGGCTGTGCATCCCTGCATCACCCCTGTTTCAGCGCCAGC 260
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119427 TCGGTGCCACGCTCTGGCACTCGCGCGCGCGCTGCTCCAGCTCGCGCTCCAGCGCCAGC 119486
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  261 CAAGGCAGCCCGCGCTCTGCGCGCGCGCGATC-GGCTGAGCCGCGAGTGGGGAGCAAGA 319
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 119487 ATCTGCGGTGCGACCTCTCTGCGAGTTTGGGCTGTGGCGCGCGCTCGGGCCAGCGCCA 119546
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  320 TCATTGCTAATCTAGTGAATCAGCTG 345
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119547 CCGTTGCCCTGCTGAAGCATCACTG 119572
```

Search completed: April 2, 2000, 21:42:57
Job time: 10416 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2000, 22:08:30 ; Search time 66.13 seconds
(without alignments)
1075.390 Million cell updates/sec

Title: US-08-988-242-1_COPY_1232_1825

Perfect score: 594

Sequence: 1 CAGGTACAGCGTAACGGCTT.....AAACGAGCTCCGTGGCGATA 594

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PTUS9_COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	594	100.0	3402	US-08-480-917-1	Sequence 1, Appli
2	39.2	6.6	2790	US-08-718-661-1	Sequence 1, Appli
3	38.2	6.4	2943	US-08-042-747A-7	Sequence 7, Appli
4	37.6	6.3	2415	US-08-474-379C-60	Sequence 60, Appli
5	36.6	6.2	913	US-08-217-327-3	Sequence 3, Appli
6	36.6	6.2	913	US-07-885-970A-3	Sequence 3, Appli
7	36.6	6.2	913	US-08-298-687A-3	Sequence 3, Appli
8	36.6	6.2	913	US-08-530-797-2	Sequence 3, Appli
9	36.6	6.2	913	US-08-298-829-3	Sequence 3, Appli
10	36.6	6.2	913	US-08-787-335-2	Sequence 2, Appli
11	36.6	6.2	1984	US-07-885-970A-25	Sequence 25, Appli
12	36.6	6.2	1985	US-08-298-687A-25	Sequence 25, Appli
13	36.6	6.2	1985	US-08-298-829-25	Sequence 25, Appli
14	36	6.1	5452	US-09-130-114-1	Sequence 1, Appli
15	36	6.1	10596	US-07-884-811-15	Sequence 15, Appli
16	36	6.1	10596	US-07-885-971-15	Sequence 15, Appli
17	36	6.1	10596	US-08-087-783A-15	Sequence 15, Appli
18	36	6.1	10596	US-08-194-088B-15	Sequence 15, Appli
19	36	6.1	10596	US-08-194-087-15	Sequence 15, Appli
20	36	6.1	10596	PCT-US93-04648-15	Sequence 15, Appli
21	35.2	5.9	3376	US-08-320-559-29	Sequence 29, Appli
22	35.2	5.9	3376	PCT-US94-04496-29	Sequence 29, Appli
23	35	5.9	168	US-08-469-802B-4	Sequence 4, Appli
24	35	5.9	168	US-08-267-803B-4	Sequence 4, Appli
25	35	5.9	171	US-08-469-802B-5	Sequence 5, Appli
26	35	5.9	171	US-08-267-803B-5	Sequence 5, Appli
27	35	5.9	195	US-08-469-802B-2	Sequence 2, Appli

c 28	35	5.9	195	3	US-08-267-803B-2	Sequence 2, Appli
c 29	35	5.9	234	2	US-08-469-802B-3	Sequence 3, Appli
c 30	35	5.9	234	3	US-08-267-803B-3	Sequence 3, Appli
c 31	35	5.9	336	4	US-07-814-220-3	Sequence 3, Appli
c 32	35	5.9	336	4	US-07-812-421-3	Sequence 3, Appli
c 33	34.8	5.9	1299	1	US-07-661-610C-11	Sequence 11, Appli
c 34	34.8	5.9	2001	2	US-08-674-168-24	Sequence 24, Appli
c 35	34.8	5.9	5437	1	US-07-661-610C-1	Sequence 1, Appli
c 36	34.8	5.9	5437	4	US-08-146-930-1	Sequence 1, Appli
c 37	34.8	5.9	6530	4	PCT-US93-03993-1	Sequence 1, Appli
c 38	34	5.7	1769	4	US-08-765-268A-1	Sequence 1, Appli
c 39	33.4	5.6	336	4	US-07-814-220-4	Sequence 4, Appli
c 40	33.4	5.6	336	4	US-07-812-421-4	Sequence 4, Appli
c 41	33.4	5.6	3073	3	US-08-714-677-11	Sequence 11, Appli
c 42	33.4	5.6	3073	3	US-08-393-540-11	Sequence 11, Appli
c 43	33.4	5.6	3073	3	US-08-714-537-11	Sequence 11, Appli
c 44	33.2	5.6	228	7	5273901-10	Patent No. 5273901
c 45	33.2	5.6	228	7	5482709-9	Patent No. 5482709

ALIGNMENTS

RESULT 1
US-08-480-917-1
; Sequence 1, Application US/08480917
; Patent No. 5820864
; GENERAL INFORMATION:
; APPLICANT: PARAMHOS-BACCALA, Glaucia
; APPLICANT: LESENECHAL, Mylene
; APPLICANT: JOLIVET, Michel
; TITLE OF INVENTION: NEW TRYPANOSOMA CRUZI ANTIGEN, AND GENE
; TITLE OF INVENTION: ENCODING THE LATTER; THEIR APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: CHAGAS DISEASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480.917
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36400
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-480-917-1

Query Match 100.0%; Score 594; DB 2; Length 3402;
Best Local Similarity 100.0%; Pred. No. 7.1e-153;
Matches 594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGGTACAGCGTAACGGCTTTTGGCTTCATCGTACAGCGAGGTAGTGGCTCTGGCT 60

Db 1232 CAGGTACAGCGTAAACGGCTTTTCTTCAATCGTACAGCCGAGCGTACGCTGCGCTCGGCT 1291

QY 61 GACATGTCGATTTCGATTGACGATCTCCATCTCCGGTCTCCCGCAGGGAAGAACAGCAG 120

Db 1292 GACATGTCGATTTCGATTGACGATCTCCATCTCCGGTCTCCCGCAGGGAAGAACAGCAG 1351

QY 121 CCAGGCCAAAAACATCGGTAGTGGCAGCGGCAAAACCGGGTGTGTCTTCGCGCACT 180

Db 1352 CCAGGCCAAAAACATCGGTAGTGGCAGCGGCAAAACCGGGTGTGTCTTCGCGCACT 1411

QY 181 GAGCGCGGAGTAGCAGTATACCAATACGACTTCTGCGGCTGCTGCATCCCTCGATCA 240

Db 1412 GAGCGCGGAGTAGCAGTATACCAATACGACTTCTGCGGCTGCTGCATCCCTCGATCA 1471

QY 241 CCCCTCTTTACGGCCAGCCAGGCGCGGCTCTCCCGCGCGCGATCGGCTGAG 300

Db 1472 CCCCTCTTTACGGCCAGCCAGGCGCGGCTCTCCCGCGCGCGATCGGCTGAG 1531

QY 301 CCGCACGTGGGGAGCAAGATCATTTGCTTAATCTAGTGAATCAGCTGGGGGATTAAATGTCACC 360

Db 1532 CCGCACGTGGGGAGCAAGATCATTTGCTTAATCTAGTGAATCAGCTGGGGGATTAAATGTCACC 1591

QY 361 CAAAGGAGCGTCTGACGACTGAGCGCGCGGCCACGACGAGGTCTACGCGGTGACGTCC 420

Db 1592 CAAAGGAGCGTCTGACGACTGAGCGCGCGGCCACGACGAGGTCTACGCGGTGACGTCC 1651

QY 421 AGCACTACCGCCCGGAGCGCAAACTCCATAGCGGCGCAATGGCGGACCTGTGAGCGCT 480

Db 1652 AGCACTACCGCCCGGAGCGCAAACTCCATAGCGGCGCAATGGCGGACCTGTGAGCGCT 1711

QY 481 GGATTGTGGCAGCTAATAGTGTGGCCAGCGCGGCTCTCTCCACAGCGCGCGGAAA 540

Db 1712 GGATTGTGGCAGCTAATAGTGTGGCCAGCGCGGCTCTCTCCACAGCGCGCGGAAA 1771

QY 541 CCAACAGGAGGAAGAAAGGCGCTCCGCGGATGTGAAACAGAGTCCGTGGCGGATA 594

Db 1772 CCAACAGGAGGAAGAAAGGCGCTCCGCGGATGTGAAACAGAGTCCGTGGCGGATA 1825

RESULT 2

US-08-718-661-1

; Sequence 1, Application US/08718661

; Patent No. 5876972

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: Nucleic acid molecules coding for mammalian

; TITLE OF INVENTION: tumor suppressor proteins and methods for their isolation

; NUMBER OF SEQUENCES: 15

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/718,661

; FILING DATE:

; CLASSIFICATION: 530

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2790 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 542..2545

US-08-718-661-1

Query Match 6.6%; Score 39.2; DB 3; Length 2790;

Best Local Similarity 54.1%; Pred. No. 0.059;

Qy 258 AGCCAAGGACGCCGGCCTCCTGC CGCGGGCGCATCGGCTGACCCG C 304
 | | | | ||||| | | | | |
Db 380 ACCTCCAGTTC TCCACC TCCTGC CA CTCTCC ACCAGCTTCTCCAC 426

```

RESULT      6
US-07-885-970A-3
; Sequence 3, Application US/07885970A
; Patent No. 5495070
; GENERAL INVENTION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701

```

Query Match	6.2%	Score 36.6;	DB 1;	Length 913;
Best Local Similarity	58.9%	Pred. No. 0.21;		
Matches 63;	Conservative	0;	Mismatches 44;	Indels 0;
Gaps				

Qy 258 AGCCAAGGAGCGCGCGCTCCTGCGCGGCGGATCGGGCTAGCCGC 304
- - - - -
Db 380 ACCTCAGTTCTCCACCTCTGCGACTCCTCCACCAAGCTTCTCCAC 426

RESULT 7
US-08-298-687A-3
; Sequence 3, Application US/08298687A
; Patent No. 5521078
; GENERAL INFORMATION:
; APPLICANT: John, Malivakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P. O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,687A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKF15A1
; CLONE: H6
; PS-08-298-687A-3

Query Match 6.2%; Score 36.6; DB 1; Length 913;
Best Local Similarity 58.9%; Pred. NO. 0.21;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY	258	AGCAAGGCAGCCGCGCCTCTGCGCGCGCGGATCGGCTGAGCCGC	304
Db	380	ACTCCAGCTTCTCCACCTCTGCCACTCTCCACAGTTCTCCAC	426

RESULT 8

GENERAL INFORMATION:
APPLICANT: John, Maliyakal E.
APPLICANT: Umbeck, Paul F.
APPLICANT: Brill, Winston J.
TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
TITLE OF INVENTION: FOR ALTERED FIBER
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Quarles and Brady
STREET: P.O BOX 2113
STREET: FIRST WISCONSIN PLAZA
CITY: MADISON
STATE: WISCONSIN
COUNTRY: U.S.A.
ZIP: 53701
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.50 inch, 800kb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/787.335
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/530,797
FILING DATE:
APPLICATION NUMBER: US 07/253,243
FILING DATE: 04-OCT-88
ATTORNEY/AGENT INFORMATION:
NAME: Nicholas J. Seay
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 1122990245
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 913 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Gossypium hirsutum
STRAIN: Coker 312
DEVELOPMENTAL STAGE: 15 day old fiber cells
TISSUE TYPE: fiber cells
IMMEDIATE SOURCE:
LIBRARY: CKFB15A1
CLONE: H6
PS-08-787-335-2

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2000, 19:13:05 ; Search time 1279.51 Seconds
(without alignments)
1752.815 Million cell updates/sec

Title: US-08-988-242-1-copy_1232_1825
Perfect score: 594
Sequence: 1 CAGGTACACGTAACGGCTT.....AAACGAGCTCCGTGGCGATA 594

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

- 1: em_est1.*
- 2: em_est2.*
- 3: em_est3.*
- 4: em_est4.*
- 5: em_est5.*
- 6: em_est6.*
- 7: em_est7.*
- 8: em_est8.*
- 9: em_est9.*
- 10: em_est10.*
- 11: em_est11.*
- 12: em_est12.*
- 13: em_est13.*
- 14: em_est14.*
- 15: em_est15.*
- 16: em_est16.*
- 17: em_est17.*
- 18: em_est18.*
- 19: em_est19.*
- 20: gb_est1.*
- 21: gb_est2.*
- 22: gb_est3.*
- 23: gb_est4.*
- 24: gb_est5.*
- 25: gb_est6.*
- 26: gb_est7.*
- 27: gb_est8.*
- 28: gb_est9.*
- 29: gb_est10.*
- 30: gb_est11.*
- 31: gb_est12.*
- 32: gb_est13.*
- 33: gb_est14.*
- 34: gb_est15.*
- 35: gb_est16.*
- 36: gb_est17.*
- 37: gb_est18.*
- 38: gb_est19.*
- 39: gb_est20.*
- 40: gb_est21.*
- 41: gb_est22.*
- 42: gb_est23.*
- 43: gb_est24.*
- 44: gb_est25.*

- 45: gb_est26.*
- 46: gb_est27.*
- 47: gb_est28.*
- 48: gb_est29.*
- 49: gb_est30.*
- 50: gb_est31.*
- 51: gb_est32.*
- 52: em_est20.*
- 53: em_est21.*
- 54: em_est22.*
- 55: em_est23.*
- 56: em_est24.*
- 57: em_est25.*
- 58: em_est26.*
- 59: gb_est33.*
- 60: gb_est34.*
- 61: gb_est35.*
- 62: gb_est36.*
- 63: gb_est37.*
- 64: gb_est38.*
- 65: em_est27.*
- 66: em_est28.*
- 67: em_est29.*
- 68: em_est30.*
- 69: gb_est39.*
- 70: gb_est40.*
- 71: gb_est41.*
- 72: gb_est42.*
- 73: gb_est43.*
- 74: gb_est44.*
- 75: em_est31.*
- 76: em_est32.*
- 77: em_est33.*
- 78: em_est34.*
- 79: gb_gss1.*
- 80: gb_gss2.*
- 81: gb_gss3.*
- 82: gb_gss4.*
- 83: em_gss1.*
- 84: em_gss2.*
- 85: em_gss3.*
- 86: em_gss4.*
- 87: gb_gss5.*
- 88: gb_gss6.*
- 89: gb_gss7.*
- 90: gb_gss8.*
- 91: gb_gss9.*
- 92: em_gss5.*
- 93: em_gss6.*
- 94: em_gss7.*
- 95: em_gss8.*
- 96: em_gss9.*
- 97: em_gss10.*
- 98: em_gss11.*
- 99: gb_gss10.*
- 100: gb_gss11.*
- 101: em_gss12.*
- 102: gb_gss12.*
- 103: gb_gss13.*
- 104: gb_gss14.*
- 105: gb_gss15.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	47.8	8.0	925	CNS0091P	AL053013 Drosophil
2	43.8	7.4	1036	CNS010BS	AL098770 Drosophil

```

3 43.4 7.3 1065 79 CNS0108S
4 41.2 6.9 427 45 A858891
5 40.4 6.8 925 79 CNS0091P
6 40.2 6.8 550 49 A0507847
7 38.8 6.5 162 79 FR0009521
8 38.8 6.5 623 33 AA413349
9 38.4 6.5 472 72 AW162815
10 38.4 6.5 472 72 AW163342
11 38.4 6.5 571 44 A1260212
12 38.4 6.5 1101 80 CNS0175Y
13 38.2 6.4 538 42 A1114969
14 38.2 6.4 814 42 A111654
15 38 6.4 436 38 A756319
16 38 6.4 553 51 A1756282
17 38 6.4 838 42 A1159665
18 38 6.4 863 42 A1159667
19 37.6 6.3 464 64 AW089826
20 37.4 6.3 964 79 CNS003WG
21 37.2 6.3 304 62 A1920772
22 37 6.2 529 42 A1159402
23 37 6.2 602 34 AA524163
24 37 6.2 705 42 A1110630
25 37 6.2 705 51 AF063504
26 37 6.2 947 79 CNS008C0
27 37 6.2 1200 80 CNS01671
28 36.8 6.2 596 79 FR0019514
29 36.8 6.2 611 43 A03030115
30 36.8 6.2 637 43 A03031310
31 36.8 6.2 720 79 CNS00Y22
32 36.6 6.2 439 47 A1492967
33 36.6 6.2 520 51 A1727579
34 36.6 6.2 542 51 A1730787
35 36.6 6.2 543 51 A1730078
36 36.6 6.2 546 51 A1726834
37 36.6 6.2 556 51 A1729927
38 36.6 6.2 563 70 AW146257
39 36.6 6.2 611 51 A1731740
40 36.6 6.2 613 51 A1727110
41 36.6 6.2 615 51 A1730728
42 36.6 6.2 617 51 A1730726
43 36.6 6.2 625 51 A1727201
44 36.6 6.2 626 51 A1727196
45 36.6 6.2 637 51 A1728230

```

ALIGNMENTS

```

RESULT 1
CNS0091P/c 925 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila

```

melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammor in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

```

source
1..925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR19D16"
/note="end : TET3"

```

```

BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN

```

```

Query Match 8.0%; Score 47.8; DB 79; Length 925;
Best Local Similarity 11.8%; Pred. No. 0.32;
Matches 37; Conservative 151; Mismatches 125; Indels 0; Gaps 0;

```

```

QY 256 CCAGCCAAAGCGAGCGCGCTCTGCCGCGCGCGGATCGGCTGAGCCGACGTGGGAGC 315
DB 914 SBBSCSSSSMTSSSSBSCSSBSSSTSSMSBSSSSSSSSSSSSSSSSSSSSSSACVC 855
QY 316 AGATCATTTGCTAATCTAGTGAATCAGCTGGGGATTAATGCACCAAGAGCGTCGTC 375
DB 854 NASSCGCCCGGABCMSCSSSSCCGSASAGVKVRASGGAGKRGSGGGASASHSSSS 795
QY 376 AGACTGGAGCGCGCCACGACGAGGTCTACGGCGGTGACGTCCACGACTACCGCCCG 435
DB 794 ACBSSSSSCSACWSASSSSSSSSSSRSRSGGAGGSGSSSSSSSSSSSSSSASAGSVSSAS 735
QY 436 CAGCGNACAAGTCCATACCGGCGACAATGCCACCTGTGACGCTGGATTGGTGGAGCT 495
DB 734 SSSSSSSSVSSVSSVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSA 675
QY 496 AATAGTGGTGGCAGCGCGCGCTCTCTCCACGCGCGCGGAGAAACACAGAGAAGAA 555
DB 674 SMSAARSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 615
QY 556 AAGGCTCCGCGG 568
DB 614 SASSGMSSSVSS 602

```

RESULT 2

```

CNS010BS 1036 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN03H11 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL098770
VERSION AL098770.1 GI:5610381
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1036)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC

```


Fax: 0298-38-7468

Email: tsasaki@abr.affrc.go.jp

PROJECT = "RGP"

FEATURES

Location/Qualifiers

1..550

/organism="Oryza sativa"

/strain="Nipponbare"

/db_xref="taxon:4530"

/clone="S21840_22"

/clone_lib="Oryza sativa mature leaf Nipponbare"

/tissue_type="mature leaf"

124 a 120 c 201 g 103 t 2 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 6.8%; Score 40.2; DB 49; Length 550;

Matches 84; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 148 ACGGGAACCGGGTGTGTCTCGGCACTGACGGCGGAGTAGCAGTCATACCAAT 207

Db 309 ACGGTTTACCGGCTCAAGTCCGACACCGACAGCCCTTCTCCCTTCGACGACGAG 250

Qy 208 AGCACTTTCGCGCTGCTGATCCCTGTCATCACCCCTGTTTCAGCGCCAGCAAGCA 267

Db 249 ACCCTTGTCTGCTGCTGATCCCGACAGCTAGCCCGCATCCGCGCCAGCTCCGACC 190

Qy 268 GCGCGCTCTCGCGCGCGGCGATCGGCTGAGCCGC 304

Db 189 GCAGGCGCTGTTCTCCCTCGCCAGCGAGCGCTCCTC 153

RESULT 7

FR0009521

LOCUS

F.rubripes GSS sequence, clone 021G08aA1, genomic survey sequence.

ACCESSION AL000805

VERSION AL000805.1 GI:2438657

KEYWORDS GSS; genome survey sequence.

SOURCE Fugu rubripes.

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;

Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;

Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.

REFERENCE 1 (bases 1 to 162)

Elgar, S., Clark, M., Smith, S., Meek, S., Warner, S., Umranta, Y.,

Williams, G. and Brenner, S.

Direct Submission

Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource

Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk

V_type: phagemid

PRIMER: KS

DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic

sequence.

FEATURES

source

1..162

/organism="Fugu rubripes"

/db_xref="taxon:31033"

/clone_lib="cosmid 021G08"

/clone="021G08aA1"

BASE COUNT 8 a 54 c 43 g 50 t 7 others

ORIGIN

Query Match

Best Local Similarity 6.5%; Score 38.8; DB 79; Length 162;

Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 170 CCTCGGCACTGACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCGCTGCTGCAT 229

Db 3 CTNCGCGNCTGCTGNGCTACTNCTGCTGCTACTACTGCTGCTGCTGCTGCTGCTA 62

Qy 230 CCCCTGCATCACCCCTGTTTTCAGCGCCAGCAAGGAGCGCGCTCTCTGCGCGCGC 289

Db 63 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122

Qy 290 GATCGGCTG 298

Db 123 CTGCTGCTG 131

RESULT 8

AA413349 623 bp mRNA EST 02-MAY-1997

LOCUS AGEST00022 Anopheles gambiae adult pSport cDNA Anopheles gambiae

DEFINITION cDNA clone cc29 5', mRNA sequence.

ACCESSION AA413349

VERSION AA413349.1 GI:2071903

KEYWORDS EST.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;

Culicoidae; Culicidae; Anopheles.

REFERENCE 1 (bases 1 to 623)

AUTHORS Cornet, A.J., Kumar, V., Mukabayire, O., Salazar Rafferty, C.,

Petrarca, V., Coluzzi, M. and Collins, F.H.

TITLE A comprehensive physical map of the malaria vector Anopheles gambiae

JOURNAL Unpublished (1997)

COMMENT On May 5, 1995 this sequence version replaced gi:797752.

Contact: Salazar Rafferty, C.; and Collins, F.H

Vector Genetics Section

Centers for Disease Control and Prevention

MS F-22, 4770 Buford Hwy, Chamblee, GA 30341

Tel: 770 488 7463

Fax: 770 488 7469

Email: czs7@cdc.gov

Mapping location: 23C in 2L chromosome

Seq primer: M13 Reverse

High quality sequence stop: 623.

FEATURES

source

1..623

/organism="Anopheles gambiae"

/strain="G3"

/db_xref="taxon:7165"

/clone="cc29"

/clone_lib="Anopheles gambiae adult pSport cDNA"

/sex="Male and Female"

/note="Site 1: SalI; Site 2: NotI; See: Salazar, C.E., et al. Insect Molecular Biology (1994), 3:1-13."

BASE COUNT 154 a 196 c 221 g 52 t

ORIGIN

Query Match 6.5%; Score 38.8; DB 33; Length 623;

Best Local Similarity 50.5%; Pred. No. 28;

Matches 94; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 102 CCGCAGGGAAGAACAGCAGCCAGCCAAAACATCGGTAGTGGGACGCGGAAACCGG 161

Db 258 CCGGCGCGCGGACAGCAGCGCTCCGAGAGAGGACCGCGCTGCGCGCGCTGC 317

Qy 162 GTGTGTCTCGGCACTGACGCGGCGAGTAGCAGTCATACCAATACGACTTCTGCCGC 221

Db 318 TGCTGAACCAAGCAAGCCTGCGGTGAGCGCAAGAGGATGCCAAGAGCTGCCCGGC 377

Qy 222 TGCTGCATCCCTGTCATCACCCCTGTTTCAGCGCCAGCCAGGAGCGCGCTCTCTGC 281

Db 378 CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 437

Qy 282 CGCGGC 287

Db 438 TGCTGC 443


```
RESULT 11
AI260212/c
LOCUS
DEFINITION
  AI260212 571 bp mRNA EST 17-NOV-1998
  LP03924.5 Prime LP Drosophila melanogaster larval-early pupal pot2
  Drosophila melanogaster cDNA clone LP03924.5 prime similar to
  X61945: n3 FBgn0010295 PID: g296042 SWISS-PROT: P40140, mRNA
  sequence.
ACCESSION
AI260212
VERSION
AI260212.1 GI: 3867737
KEYWORDS
EST.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 571)
  Harvey, D., Hong, L., Evans-Hoim, M., Pendleton, J., Su, C.,
  Brokstein, P., Lewis, S. and Rubin, G.M.
  BDCP/HMI Drosophila EST Project
  Unpublished (1997)
  Contact: Harvey, D.
  G. M. Rubin-Molecular and Cell Biology
  University of California Berkeley
  539 LSA, Berkeley, CA 94720-3200, USA
  Fax: 510 643 9947
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  Sequence is the complete cDNA insert. hit genomic sequence AL024484
  Plate: 39 row: B column: 12
  High quality sequence stop: 513.
  Location/Qualifiers
    1..571
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone_lib="LP03924"
      /sex="male and female"
      /dev_stage="larvae-pupae"
      /lab_host="DH5-alpha"
      /note="Organ: whole body; Vector: pot2; Site_1: EcoRI;
      Site_2: XhoI; Sized fractionated cDNAs were directly
      ligated into pot2. plasmid cDNA library."
BASE COUNT
161 a 169 c 114 g 126 t 1 others
ORIGIN
Query Match 6.5%; Score 38.4; DB 44; Length 571;
Best Local Similarity 56.2%; Pred. No. 34;
Matches 72; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 154 AACCGGGGTGTCTCTCGGCGACTACGCGCGGAGTAGCAGTCATACCAATACGACT 213
  |||||
DB 427 AGACAGCGCTCGGCACATCGGCTCTCTCTCGGCGGATGCGACGTTCACCACTCGGCT 368
  |||||
QY 214 TCTGCGCTGTGATCCCTGCATCACCCCTGTTCAGGCCAGCAAGCAGCCGCG 273
  |||||
DB 367 ATGTGTGTCACACTGCTGCTACTGCTACTGCTGCTGCGGCGACGACGCGCTCGTGT 308
  |||||
QY 274 CTTCTCTGC 281
  |||||
DB 307 GCTGGTGC 300

RESULT 12
CNS017SY
LOCUS
DEFINITION
  CNS017SY 1101 bp DNA GSS 26-JUL-1999
  Drosophila melanogaster genome survey sequence SP6 end of BAC
  BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
AL108460
VERSION
AL108460.1 GI: 5628764
```

```
KEYWORDS
fruit fly.
SOURCE
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
  Genoscope.
  Direct Submission
  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
  - Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk -- This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaut at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBelOBAC11.
  Location/Qualifiers
    1..1101
      /organism="Drosophila melanogaster"
      /plasmid="pBelOBAC11"
      /db_xref="taxon:7227"
      /clone_lib="DrosBAC"
      /clone="BACN37L08"
      /note="end : SP6"
BASE COUNT
254 a 176 c 160 g 152 t 359 others
ORIGIN
Query Match 6.5%; Score 38.4; DB 80; Length 1101;
Best Local Similarity 15.1%; Pred. No. 40;
Matches 67; Conservative 182; Mismatches 195; Indels 0; Gaps 0;
QY 137 CGGTAGTCGCGACGCGGAAACCGGGTGTCTCTCGGCGACTACGCGCGGATGCA 196
  |||||
DB 638 CGGAGCGGRKGCACAKSAGMSGCGRSGSGGSCCGKAKGVRGVRVCCAGGGASC 697
  |||||
QY 197 GTCATCAATACGACTTCTCCGCTGCTGCATCCCTGCATCACCCCTCTTTTCAGCGC 256
  |||||
DB 698 ACMAADCGCCAKMACCCSSSSASGSSCASTSSASRGMVSSCACSGSGGASACGA 757
  |||||
QY 257 CAGCCAGGCGCGCGCTCTCTCGCGCGCGCGATCGGCTGAGCCGACGTGGGAGCA 316
  |||||
DB 758 SGAGGSGRGGGCGCCASSGVCGAACSSASCSASGSGGSSCSASCGSCCGVSSCSA 817
  |||||
QY 317 AGATCATTTGCTAATCTAGTGAATCAGCTGGGATTAATGTACCCAAAGGAGCGTCGTA 376
  |||||
DB 818 VSASASVMSKVASAVASCSAVASGMSAGAVSSCRSSVASSVSSSSSSSSSSSS 877
  |||||
QY 377 GCACCTGGAGCGCGCGCACGACGAGGTCTACGGCGGTGACGTCCACGACTACCGCCCGC 436
  |||||
DB 878 SVVSAVAVSSSSSASSASMAVAVAVAVAVSSVSVSSSSSSSSSSSSSVSVASVA 937
  |||||
QY 437 AGCGAACAGTCCATACGCGCACATGCCCGACCTGTGACGCGTGGATTGGTGGCAGTA 496
  |||||
DB 938 ASASVSSSSSSSVSSSVSSSVASVMSVAVSSSSSSSSSSSVSSSVSVVAASAAAA 997
  |||||
QY 497 ATAGTGTGCGAGCGCGCTCTCTCCACAGCGCGCGCAACCAACAGGAGAGAA 556
  |||||
DB 998 AAAAAASSSASAVAVVSVSSSSSSSSSSSSSSSSSSSSSVSSSVSVSSSVSV 1057
  |||||
QY 557 AGGCTCCGCGCATGTGAACGA 580
  |||||
DB 1058 AVASASASVSSVAAVAVAAMAA 1081

RESULT 13
AI114969/c
LOCUS
DEFINITION
  AI114969 538 bp mRNA EST 02-SEP-1998
  u141b06.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
```

IMAGE:1884947 5' similar to TR:014633 014633 SKIN-SPECIFIC PROTEIN.
 ; mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

A1114969
 A1114969.1 GI:3515293
 EST.
 house mouse.

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 538)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE
 JOURNAL
 COMMENT

The WashU-HMI Mouse EST Project
 Unpublished (1996)
 On Jan 17, 1998 this sequence version replaced gi:1900380.

CONTACT
 WASHINGTON UNIVERSITY SCHOOL OF MEDICINE
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:969271

Seq primer: custom primer used
 High quality sequence stop: 490.
 Location/Qualifiers

1 538
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:1884947"
 /dev_stage="embryo, 14 dpc"
 /lab_host="DH10B"
 /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTC);
 Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
 with an oligo(dt) primer [ATGTGGCCTTTTITTTTTTTT];
 double-stranded cDNA was ligated to a DraIII adaptor
 [TGTGGCCTACTGG], digested and cloned into distinct DraIII
 sites of the pME18S-FL3 vector (5' site CACTGTGTC, 3' site
 CACCATGTG). XhoI should be used to isolate the cDNA
 insert. Size selection was performed to exclude fragments
 <1.5kb. Library constructed by Dr. Sumio Sugano
 (University of Tokyo Institute of Medical Science).
 Custom primers for sequencing: 5' end primer
 CTTCTGCTCTAAAGCTGGC and 3' end primer
 CGACCTGAGCTCGAGCACA."

BASE COUNT
 ORIGIN

91 a 149 c 190 g 108 t
 Query Match 6.4%; Score 38.2; DB 42; Length 538;
 Best Local Similarity 52.9%; Pred. No. 37;
 Matches 82; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 133 ACATCGGTAGTGGCAGCGGAAACCGGGTGTGTCTCGGGCACTGACGGCGGAGT 192

DB 439 ACCACTGTCTGCCACAGCAACTGTGCCACCGCTGTCTGCCACAGCAACTGTGCC 380

QY 193 AGCAGTCATACCAATACGACTTCTGCCGCTGTGCATCCCTCCATCACCCCTGTTTCA 252

DB 379 ACCGCTGTCTGCCACAGCAACTGTGCCACCGCTGTGCCACCGCTGTGCCACCGTGTCT 320

QY 253 GCGCCAGCCAGCAGCGCGCTCTGCCCGGC 287

DB 319 GCCACAGCAACTGTGCCACCGCTGTGCCACAGC 285

RESULT 14
 A1131654

LOCUS
 DEFINITION

A1131654 814 bp mRNA EST 14-SEP-1998
 ui41b06.x1 Sugano mouse embryo mewa Mus musculus cDNA clone
 IMAGE:1884947 3' similar to TR:014633 014633 SKIN-SPECIFIC PROTEIN.
 ; mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE

A1131654
 A1131654.1 GI:3601670
 EST.
 house mouse.

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 814)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

REFERENCE
 AUTHORS

The WashU-HMI Mouse EST Project
 Unpublished (1996)
 On Sep 12, 1996 this sequence version replaced gi:1402281.
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu

TITLE
 JOURNAL
 COMMENT

On Sep 12, 1996 this sequence version replaced gi:1402281.
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:969271

Seq primer: custom primer used
 High quality sequence stop: 375.
 Location/Qualifiers

1 814
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:1884947"
 /dev_stage="embryo, 14 dpc"
 /lab_host="DH10B"
 /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTC);
 Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
 with an oligo(dt) primer [ATGTGGCCTTTTITTTTTTTT];
 double-stranded cDNA was ligated to a DraIII adaptor
 [TGTGGCCTACTGG], digested and cloned into distinct DraIII
 sites of the pME18S-FL3 vector (5' site CACTGTGTC, 3' site
 CACCATGTG). XhoI should be used to isolate the cDNA
 insert. Size selection was performed to exclude fragments
 <1.5kb. Library constructed by Dr. Sumio Sugano
 (University of Tokyo Institute of Medical Science).
 Custom primers for sequencing: 5' end primer
 CTTCTGCTCTAAAGCTGGC and 3' end primer
 CGACCTGAGCTCGAGCACA."

BASE COUNT
 ORIGIN

184 a 245 c 211 g 173 t 1 others
 Query Match 6.4%; Score 38.2; DB 42; Length 814;
 Best Local Similarity 52.9%; Pred. No. 41;
 Matches 82; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 133 ACATCGGTAGTGGCAGCGGAAACCGGGTGTGTCTCGGGCACTGACGGCGGAGT 192

DB 423 ACCACTGTCTGCCACAGCAACTGTGCCACCGCTGTCTGCCACAGCAACTGTGCC 482

QY 193 AGCAGTCATACCAATACGACTTCTGCCGCTGTGCATCCCTCCATCACCCCTGTTTCA 252

DB 483 ACCGCTGTCTGCCACAGCAACTGTGCCACCGCTGTGCCACCGCTGTGCCACCGTGTCT 542

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 2, 2000, 21:55:01 ; Search time 102.54 Seconds
(without alignments)
1449.328 Million cell updates/sec

Title: US-08-988-242-1_COPY_1232_1825
Perfect score: 594
Sequence: 1 CAGGTACAGCGTACGGCCTT.....AAACGAGCTCGTGGCGATA 594

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	590.8	99.5	3402	1 T27310	Trypanosoma cruzi
2	39.2	6.6	2790	1 V18480	BOP1 CDNA, New Iso
3	38.2	6.4	2943	1 T16480	SAB virus gB glyco
4	36.8	6.2	3776	1 T42903	TRP-1 protein codi
5	36.6	6.2	203	1 V17226	SCA2 gene CAG repe
6	36.6	6.2	203	1 V30271	Glutamine rich reg
7	36.6	6.2	913	1 T07199	Cotton fibre-speci
8	36.6	6.2	913	1 T13034	Cotton fibre-speci
9	36.6	6.2	913	1 T30253	Cotton fibre cell-
10	36.6	6.2	913	1 T62610	Cotton fibre speci
11	36.6	6.2	913	1 T70041	Cotton fibre speci
12	36.6	6.2	1884	1 T13030	Cotton fibre-speci
13	36.6	6.2	1985	1 T30250	Cotton fibre-speci
14	36.6	6.2	1985	1 T70036	Cotton fibre clone
15	36.4	6.1	2415	1 T86757	CDNA of the M3/6 g
16	36.4	6.1	2453	1 T86758	CDNA of the M3/6 g
17	36.2	6.1	633	1 Q03322	Genomic Eimeria te
18	36	6.1	795	1 V55830	FLGA insert stabl
19	36	6.1	799	1 V55831	Nucleotide sequenc
20	36	6.1	9600	1 V21683	Vector plasmid pCM
21	36	6.1	10596	1 Q51731	Plasmid pcisEBON f
22	36	6.1	10596	1 T40348	Plasmid pcisEBON f
23	36	6.1	10596	1 X15650	Nucleotide sequenc
24	35.8	6.0	3076	1 V43674	Receptor type tyro
25	35.6	6.0	1089	1 N92576	Sequence of the 1.
26	35.6	6.0	1089	1 T93593	Eimeria tenella sp
27	35.6	6.0	114955	1 X53491	Human adenosine Al
28	35.4	6.0	1463	1 V61483	Human secreted pro
29	35.2	5.9	417	1 Q25082	Antigen tc-11e gen
30	35.2	5.9	3376	1 Q75166	AF-9 CDNA, New acu
31	35	5.9	168	1 Q84833	Spinocerebellar at
32	35	5.9	171	1 Q84834	Spinocerebellar at
33	35	5.9	195	1 Q84831	Spinocerebellar at
34	35	5.9	234	1 Q84832	Spinocerebellar at

c 35 35 5.9 292 1 T75505 P. americanus skin
c 36 35 5.9 627 1 V90494 EST clone DN740. N
c 37 34.8 5.9 2001 1 T49282 DNA encoding B. la
c 38 34.8 5.9 2001 1 T84788 Brevibacterium lac
c 39 34.8 5.9 2001 1 V15788 B. lactofermentum
c 40 34.8 5.9 2001 1 V40257 Brevibacterium lac
c 41 34.8 5.9 6530 1 Q51557 Loridin gene, Con
c 42 34.6 5.8 296 1 V89532 EST clone CP328. N
c 43 34.6 5.8 1325 1 V80740 Human secreted pro
c 44 34.6 5.8 2400 1 T42859 Choline oxidase ge
c 45 34.6 5.8 2400 1 T75000 Choline oxidase ge

ALIGNMENTS

RESULT 1

T27310
ID T27310 standard; cDNA; 3402 BP.
AC T27310;
DT 26-NOV-1996 (first entry)
DE Trypanosoma cruzi epimastigotic Ptc100t antigen gene.
KW Antigen; Trypanosoma cruzi; epimastigote; serum; Chagas disease; probe;
KW Primer; PCR; polymerase chain reaction; amplification; antibody; ds.
QS Trypanosoma cruzi.
FH Key Location/Qualifiers
FT cds 266..3013
FT /*tag= a
FT /product= PTC100t epimastigotic antigen

FR2723589-A1.
16-FEB-1996.
12-AUG-1994; 010132.
12-AUG-1994; FR-010132.
PA (INMR) BIO MERIEUX.
PI Jolivet M, Lesenechal M, Paranhos-Baccala G;
DR WPI 96-190287/20.
DR P-PSDB R91615.
PT New nucleic acid encoding Trypanosoma cruzi epimastigotic antigen -
PT useful for diagnosis, monitoring and therapy of Chagas disease
PS Claim 1; Page 24-26; 53pp; French.

CC This is the nucleotide sequence encoding a novel isolated antigenic
CC protein from Trypanosoma cruzi epimastigotes, designated Ptc100t.
CC The clone Tc50 was isolated from a T. cruzi genomic expression library in
CC lambda gt11, using a mixture of sera from patients with Chagas disease
CC Clone Tc50 contained an 594 bp insert corresp. to nucleotides 1232-1825
CC of this sequence. The Tc50 sequence was subsequently used to probe a
CC Southern blot of restriction enzyme digested T. cruzi DNA and also screen
CC a lambda gt10 library to isolate a 1041 bp EcoRI fragment corresp. to
CC nucleotides 1403-2443 of Ptc100t. Primers (T27311-5) were synthesised
CC based on the sequences of the 594 and 1041 bp fragments and used to
CC amplify the Ptc100t clone as 3 fragments from cDNA derived from mRNA
CC purified from T. cruzi epimastigotes. The protein or antibodies raised
CC against it can be used in the detection and monitoring of T. cruzi
CC infection i.e. Chagas disease.
SQ Sequence 3402 BP; 889 A; 818 C; 958 G; 737 T;

Query Match 99.5%; Score 590.8; DB 1; Length 3402;
Best Local Similarity 99.7%; Pred. No. 1.2e-140;
Matches 592; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGGTACAGCGTAAAGCGCTTTTGGCTTCAATCGTACAGCGAGGTAGCTCGCTCGGCT 60

DB 1232 CAGGTACAGCGTAAAGCGCTTTTGGCTTCAATCGTACAGCGAGGTAGCTCGCTCGGCT 1291

QY 61 GACATGTCGATTCGATTCGATTCCTCCGTCCTCCGTCCTCCGTCCTCCGTCCTCCGTCCT 120

DB 1292 GACATGTCGATTCGATTCGATTCCTCCGTCCTCCGTCCTCCGTCCTCCGTCCTCCGTCCT 1351

QY 121 CCAGGCCAAAACATCGGTAGTGGCGAGCGGGAACACCGGGGTGTGTCTCGGGCACT 180

DB 1352 CCAGGCCAAAACATCGGTAGTGGCGAGCGGGAACACCGGGGTGTGTCTCGGGCACT 1411

QY	181	GAGCGGCGAGTAGCAGTATACCAATACGACTTCTGCCGTGTGTCATCCCTGTCATCA	240
Db	1412	GAGCGGCGAGTAGCAGTATACCAATACGACTTCTGCCGTGTGTCATCCCTGTCATCA	1471
QY	241	CCCCCTGTTTACGCGCCAGCCAGGACGCGCGCTCTCGCGCGGCGGATCGGCTGAG	300
Db	1472	CCCCCTGTTTACGCGCCAGGAGGACGCGCGCTCTCGCGCGGCGGATCGGCTGAG	1531
QY	301	CCGACGCTGGGGAGCAAGATCATTTGCTAATCTAGTGAATACGCTGGGGATTAATGTCA	360
Db	1532	CCGACGCTGGGGAGCAAGATCATTTGCTAATCTAGTGAATACGCTGGGGATTAATGTCA	1591
QY	361	CAAGAGGAGCGTCTGTCAGCACTGGAGCGCGGCCACGACGAGGTCTACGGCGGTGACGTCC	420
Db	1592	CAAGAGGAGCGTCTGTCAGCACTGGAGCGCGGCCACGACGAGGTCTACGGCGGTGACGTCC	1651
QY	421	ACGACTACGCGCCCGCAGCGACAGTCCATACGCGGACAAATGCGCGACCTGTGACGGCT	480
Db	1652	ACGACTACGCGCCCGCAGCGACAGTCCATACGCGGACAAATGCGCGACCTGTGACGGCT	1711
QY	481	GGATTGGTGGCAGCTAATAGTGGTCCAGCGCGGCTCTGCTCCACAGCGCGCGGAAA	540
Db	1712	GGATTGGTGGCAGCTAATAGTGGTCCAGCGCGGCTCTGCTCCACAGCGCGCGGAAA	1771
QY	541	CCAACAGGAGAGAAAAGCGCTCCGCGGCATGTAAACGAGCTCCGTGGCGATA	594
Db	1772	CCAACAGGAGAGAAAAGCGCTCCGCGGCATGTAAACGAGCTCCGTGGCGATA	1825
RESULT 2			
ID	V18480	V18480 standard; cDNA to mRNA; 2790 BP.	
AC	V18480;		
DC	18-AUG-1998	(first entry)	
DE	BOP1 CDNA.		
KW	Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;		
KW	inhibible; alzheimer's disease; nuclear transcription factor; apoptosis		
KW	cell cycle; neuronal disorder; ss.		
OS	Mus sp.		
FH	Key	Location/Qualifiers	
FT	CDS	542..2545	/*tag= a
FT		/product= "BOP1 protein"	
FT	WO9813489-A1.		
PN	02-APR-1998.		
PD	22-SEP-1997; E05198.		
PF	23-SEP-1996; US-718661.		
PR	(CNRS) CENT NAT RECH SCI.		
PA	(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.		
PI	Journot L, Spengler D;		
DR	WPI: 98-230701/20.		
DR	P-PSDB; W48760.		
PT	New isolated tumour suppressor gene - useful for developing products		
PT	for use in diagnosis and treatment of tumour(s) or neuronal		
PT	disorder(s)		
PS	Claim 1: Pages 72-76; 118pp; English.		
CC	The present sequence represents the BOP1 CDNA isolated from the		
CC	mouse corticotroph pituitary tumour cell line Atr-20 cDNA library.		
CC	The protein encoded by the BOP1 CDNA displays a tumour suppressing		
CC	activity when it was constitutively and inducibly expressed in		
CC	tumour cells. The BOP1 CDNA and the protein it encodes are claimed		
CC	to be useful in the preparation of therapeutic compositions, useful		
CC	for treating, preventing or delaying the recurrence of a tumour or		
CC	neural disorders, e.g. genetic diseases or acquired degenerative		
CC	encephalopathies such as Alzheimer's disease. The BOP1 protein is		
CC	also claimed to be able to induce apoptosis resulting in inhibition		
CC	of tumour cell growth, to suppress tumour formation, to induce G1		
CC	arrest of the cell cycle and to act as nuclear transcription factor.		
CC	Sequence 2790 BP; 567 A; 783 C; 714 G; 626 T;		
SQ			

Query Match 6.6%; Score 39.2; DB 1; Length 2790;

Best Local Similarity 54.1%; Pred. No. 0.31;

Matches	80;	Conservative	0;	Mismatches	68;	Indels	0;	Gaps
QY	137	CGGTAGTGGCGACGGCGAAACCGGGTGTGTCTCTCGGCACACTACCGCGGAGTAGCA	196					
Db	2114	CAGCAGCAGCCACCGCCACTTGC	CGCGCCACACAGCTGCCACTGCCACGACCACTGCA	2173				
QY	137	GTCAACAATACACACTTCTGCGCGTCTGCTGCATCCCTCATCACCCCTGTTTTCAGGCC	256					
Db	2174	GTGCCACAGCCACTACCA	CAGCAGCCACAGATGCAGCCACAGATTCAGTTGCATGCATCAGGCC	2233				
QY	257	CAGCCAAAGCAGCGCGCTCTCTGCCGC	284					
Db	2234	CAGATGCAGCTACCA	CAGCTGCTGCCGC	2261				
RESULT	3							
TI6480/c								
ID	TI6480	standard; DNA; 2943 BP.						
AC	TI6480;							
DT	11-MAY-1996	(first entry)						
DE	SA8 virus gB glycoprotein coding sequence.							
KW	Herpes simian monkey SA8 virus gB glycoprotein; Immunoassay;							
KW	diagnosis; herpes B virus; ss.							
OS	Herpes simian monkey SA8 virus.							
FH	Key	Location/Qualifiers						
FT	cds	87..2744						
FT		/tag= a						
FT		/product= gB_glycoprotein						
PN	US5487969-A.							
PD	30-JAN-1996.							
PF	01-APR-1993; 042747.							
PR	01-APR-1993; US-042747.							
PA	(SWBI-) SOUTHWEST FOUND BIOMEDICAL RES.							
PI	Black D., Eberle R., Hilliard J., Scinicariello F;							
PR	WPI; 96-105220/11.							
P	P-PSDB: R92747.							
PT	Detection of herpes B virus by PCR amplification of sample DNA - to							
PT	detect a specific herpes simian monkey B virus DNA segment.							
PS	Disclosure; Column 23-30; 22pp; English.							
CC	This sequence encodes the herpes simian monkey SA8 virus gB							
CC	glycoprotein. Such sequences have been used to develop synthetic							
CC	DNA primers which have homologous sequences of conserved regions							
CC	which flank a divergent region of the gB glycoprotein gene. In							
CC	addition, proteins such as the gB glycoprotein have potential use in							
CC	the development of serological immunoassays. One approach is to							
CC	synthesize peptides which, based on the properties of the predicted							
CC	protein sequence, are likely to be immunologically active. Such							
CC	peptides can be used as substrate antigens in immunoassays to detect							
CC	serum antibodies which recognize this specific peptide sequence.							
CC	Synthetic peptides may also be used to produce antibodies against							
CC	specific regions of the gB glycoprotein which are unique to one							
CC	virus. These can then be used to develop virus-specific							
CC	immunoassays for differentiation of SA8 virus from other primate							
CC	alpha-herpes viruses and for identification of antibodies directed							
CC	against SA8 virus in primate serum samples.							
SQ	Sequence	2943 BP; 462 A; 1103 C; 983 G; 395 T;						
Query Match		6.4%; Score 38.2; DB 1; Length 2943;						
Best Local Similarity		47.7%; Pred. No. 0.57;						
Matches	112;	Conservative	0;	Mismatches	123;	Indels	0;	Gaps
QY	209	CGACTTCTGCGGTGCTGCATCCCTTGCATCACCCCTGTTTTCAGCGCCAGCAGGACG	268					
Db	255	CGTCTCTGTCGCGAGCTCTCTCTGCTCGGGAAGGCGCGGCTGCGCGGGTCTCGCGG	196					
QY	269	CCGCGCTCTCTGCGCGGCGCATCGCTTTCAGCCACAGTGGGAGCAAGATCTGCTA	328					
Db	195	TCGGGCGCGCGCGCGCGCGCGCGCTGCTGCGCGCGCGCGATCATCGCGAGGCGCA	136					
QY	329	ATCTAGTGAATCAGCTGGGGATTATGTTCACCAAGAGGAGGCTCTGCACACTGGAGCGC	388					
Db	135	GGAGACCGGAGGGAAGAAGAGGGGGGGTCTCCGAGGCGCATCTCCGGGGCGCG	76					

```
Oy 389 CGGCACGACGAGTCTACGGGGTGAGTCCACGACTACGCGCCGCGAGCAAC 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 CGGGCGCTACGGGGCGGACTACGGGGGGTGTGACTCCGGCGCGCGCGGAGC 21
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
T42903
ID T42903 standard; DNA; 3776 BP.
AC T42903;
DE 16-JUN-1997 (first entry)
KW TRP-1 protein coding sequence.
KW gene expression; regulation; plasmid; viral infection;
KW human T-cell leukaemia; HIV; antiviral agent; detection; cancer;
KW gene therapy; TSP; ds.
OS Homo sapiens.
FT key Location/Qualifiers
FT cds 139..2205
FT /*tag= a
FT /product= TRP-1
PN WO9630522-A1.
PD 03-OCT-1996.
PF 19-MAR-1996; J00719.
PR 24-MAR-1995; JP-066559.
PR 27-APR-1995; JP-104299.
PA (SHIO) SHIONOGI & CO LTD.
PI Igarashi H, Okumura K, Orita S, Saiga A, Sakaguchi G;
DR WPI; 96-455367/45.
DR P-PSDB; W06136.
PT DNA molecule with gene expression regulation activity - for use in
PT e.g. treatment of human T-cell leukaemia and HIV, as antiviral agent
PT and for detecting cancer
PS Claim 17; Page 48-54; 77pp; Japanese.
CC This sequence encodes the protein TSP-1. This sequence was used in
CC conjunction with a DNA molecule with gene expression regulation activity,
CC in a plasmid for regulation of gene expression, and treatment of viral
CC infection pref. human T-cell leukaemia and HIV. The plasmid also encodes
CC a protein which is used as an antiviral agent, and also in a method for
CC detecting cancer. The DNA molecule and protein have potential uses in
CC gene therapy, and the plasmid may also have potential use in the
CC treatment of TSP.
SQ Sequence 3776 BP; 764 A; 1177 C; 1183 G; 652 T;

Query Match 6.2%; Score 36.8; DB 1; Length 3776;
Best Local Similarity 69.4%; Pred. No. 1.4;
Matches 50; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Oy 231 CCTGTCATCACCCCTGTTTCAGCGCCAGCGCGGCTCTCTGCGCGGCGG 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1355 CCGCGCCACAGCCCGAGCCCGCCAGCCAGCCAGCGCGGCTGAGTCCGAGCC 1414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 291 ATCGGCTGAGCC 302
    ||| ||| ||| |||
Db 1415 AGCCCCAGAGCC 1426

RESULT 5
V17226/c
ID V17226 standard; DNA; 203 BP.
AC V17226;
DE 29-JUN-1998 (first entry)
DE SCA2 gene CAG repeat unit fragment.
KW SCA2 gene; spinocerebellar ataxia type II; CAG repeat; PCR primer; ss.
OS Synthetic.
PN WO9803679-A1.
PD 29-JAN-1998.
PF 18-JUL-1996; J01999.
PR 18-JUL-1996; WO-J01999.
PA (SRLS-) SRL INC.
PI Sanpei K, Tsuji S;
DR WPI; 98-120796/11.
PT Diagnosing spinocerebellar ataxia type II - by PCR and determining
```

```
PT number of CAG repeat units
PS Example 1; Page 12; 23pp; Japanese.
CC This sequence represents a fragment of the SCA2 gene. It can be used in
CC the method of the invention for diagnosing spinocerebellar ataxia type
CC II, by performing PCR on the test DNA using two primers hybridising to
CC parts of the SCA2 gene sequence, and determining the number of CAG
CC repeats in the amplified products. The method provides an easy means for
CC the diagnosis of spinocerebellar ataxia type II. 3 T;
SQ Sequence 203 BP; 68 A; 70 C; 62 G; 3 T;

Query Match 6.2%; Score 36.6; DB 1; Length 203;
Best Local Similarity 57.4%; Pred. No. 0.79;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Oy 184 GCGGCGAGTAGCAGTCATACCAATACGACTTCTGCGGCTGTGTCATCCCTGCATCACCC 243
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 202 GCCCAGAGTTTCCGTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143
    |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Oy 244 CCTGTTTCAGCGCCAGCCAGCGCGGCTCTCTGCGCGGCGGCGATCGGCTG 298
    |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 142 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 88
    |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 6
V30271/c
ID V30271 standard; DNA; 203 BP.
AC V30271;
DE 02-OCT-1998 (first entry)
DE Glutamine rich region encoding sequence found in SCA2 patients.
KW Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KW CAG repeat; neurodegenerative disease; ss.
OS Homo sapiens.
PN WO9818920-A1.
PD 07-MAY-1998.
PF 30-OCT-1997; J03946.
PR 30-OCT-1996; JP-304059.
PA (SRLS-) SRL INC.
PI Sanpei K, Tsuji S;
DR WPI; 98-27215/24.
PT Nucleic acid fragments associated with spinocerebellar ataxia type 2
PT - contain increased number of CAG repeat region compared to normal
PT gene
PS Example 1; Page 22; 38pp; Japanese.
CC This represents a sequence encoding glutamine rich repeat region which
CC can be found in spinocerebellar ataxia type 2 (SCA2) patients. The
CC specification provides a gene sequence causative of the neurodegenerative
CC disease SCA2, having a tri-nucleotide (CAG) repeat region which in the
CC expression product produces a polyglutamine sequence from Gln-166 to
CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
CC patients this number is increased to 35-100. Peptides encoded by nucleic
CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
CC gene, antibodies recognising the peptides and antisense nucleic acids
CC hybridising with the nucleic acid fragments can be used for the
CC investigation and diagnosis of SCA2. They can also be used for the
CC treatment of SCA2 by antisense therapy or gene therapy.
SQ Sequence 203 BP; 68 A; 70 C; 62 G; 3 T;

Query Match 6.2%; Score 36.6; DB 1; Length 203;
Best Local Similarity 57.4%; Pred. No. 0.79;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Oy 184 GCGGCGAGTAGCAGTCATACCAATACGACTTCTGCGGCTGTGTCATCCCTGCATCACCC 243
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||
Db 202 GCCCAGAGTTTCCGTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143
    |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Oy 244 CCTGTTTCAGCGCCAGCCAGCGCGGCTCTCTGCGCGGCGGCGATCGGCTG 298
    |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 142 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 88
    |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

RESULT 7
```



```
QY 389 CGGCCACGACGAGGTCTACGGCGGTGACGCTCCACGACTACCGCCCGCAGCGAAC 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 CGGGCCTACGGCGGCGACTACCGGGCGGTGAGTCCCGCCCGCGCGCGAGC 21
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
T42903
ID T42903 standard; DNA; 3776 BP.
AC T42903;
DE TRP-1 protein coding sequence.
KW gene expression; regulation; plasmid; viral infection;
KW human T-cell leukaemia; HIV; antiviral agent; detection; cancer;
KW gene therapy; TSP; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 139..2205
FT /tag= a
FT /product= TRP-1
FT WO9630522-Al.
PN 03-OCT-1996.
PD 19-MAR-1996. J00719.
PR 24-MAR-1995; JP-066559.
PR 27-APR-1995; JP-104299.
PA (SHIO ) SHIONOGI & CO LTD.
PI Igarashi H, Okumura K, Orita S, Saiga A, Sakaguchi G;
DR WPI; 96-453367/45.
DR P-PSDB; W06136.
PT DNA molecule with gene expression regulation activity - for use in
PT e.g. treatment of human T-cell leukaemia and HIV, as antiviral agent
PT and for detecting cancer
PS Claim 17; Page 48-54; 77pp; Japanese.
CC This sequence encodes the protein TRP-1. This sequence was used in
CC conjunction with a DNA molecule with gene expression regulation activity,
CC in a plasmid for regulation of gene expression, and treatment of viral
CC infection pref. human T-cell leukaemia and HIV. The plasmid also encodes
CC a protein which is used as an antiviral agent, and also in a method for
CC detecting cancer. The DNA molecule and protein have potential uses in
CC gene therapy, and the plasmid may also have potential use in the
CC treatment of TSP.
SQ Sequence 3776 BP; 764 A; 1177 C; 1183 G; 652 T;

Query Match 6.2%; Score 36.8; DB 1; Length 3776;
Best Local Similarity 59.4%; Pred. No. 1.4;
Matches 50; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 231 CCCTGCATCACCCCTCTTTACGGCCAGCAAGGCGCGCCCTCTCGCCGCGCGG 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1355 CCGCGCCACAGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 1414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 291 ATCGGTGAGCC 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1415 AGCCCCAGAGCC 1426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
V17226/c
ID V17226 standard; DNA; 203 BP.
AC V17226;
DE SCA2 gene CAG repeat unit fragment.
KW SCA2 gene; spinocerebellar ataxia type II; CAG repeat; PCR primer; ss.
OS Synthetic.
PN WO9803679-Al.
PD 29-JAN-1998.
PF 18-JUL-1996; J01999.
PR 18-JUL-1996; WO-J01999.
PA (SRLS-) SRL INC.
PI Sanpei K, Tsuji S;
DR WPI; 98-120796/11.
PT Diagnosing spinocerebellar ataxia type II - by PCR and determining
```

```
PT number of CAG repeat units
PS Example 1; Page 12; 23pp; Japanese.
CC This sequence represents a fragment of the SCA2 gene. It can be used in
CC the method of the invention for diagnosing spinocerebellar ataxia type
CC II, by performing PCR on the test DNA using two primers hybridising to
CC parts of the SCA2 gene sequence, and determining the number of CAG
CC repeats in the amplified products. The method provides an easy means for
CC the diagnosis of spinocerebellar ataxia type II. 3 T;
SQ Sequence 203 BP; 68 A; 70 C; 62 G;

Query Match 6.2%; Score 36.6; DB 1; Length 203;
Best Local Similarity 57.4%; Pred. No. 0.79;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 184 GCGGCGAGTAGCAGTCATACCAATACGACTTCTGCGGCTGCTGCTGATCCCTGCATCACCC 243
    || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 GCCCAGAGTTTCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143
    || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 244 CTTGTTTACGCCAGCCAGCAGCGCGCTCTCTGCGCGCGCGGCGATCGGCTG 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
V30271/c
ID V30271 standard; DNA; 203 BP.
AC V30271;
DE 02-OCT-1998 (first entry)
DE Glutamine rich region encoding sequence found in SCA2 patients.
KW Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KW CAG repeat; neurodegenerative disease; ss.
OS Homo sapiens.
PN WO9818920-Al.
PD 07-MAY-1998.
PR 30-OCT-1997; J03946.
PR 30-OCT-1996; JP-304059.
PA (SRLS-) SRL INC.
PI Sanpei K, Tsuji S;
DR WPI; 98-272215/24.
PT Nucleic acid fragments associated with spinocerebellar ataxia type 2
PT - contain increased number of CAG repeat region compared to normal
PT gene
PS Example 1; Page 22; 38pp; Japanese.
CC This represents a sequence encoding glutamine rich repeat region which
CC can be found in spinocerebellar ataxia type 2 (SCA2) patients. The
CC specification provides a gene sequence causative of the neurodegenerative
CC disease SCA2, having a tri-nucleotide (CAG) repeat region which in the
CC expression product produces a polyglutamine sequence from Gln-166 to
CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
CC patients this number is increased to 35-100. Peptides encoded by nucleic
CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
CC gene, antibodies recognising the peptides and antisense nucleic acids
CC hybridising with the nucleic acid fragments can be used for the
CC investigation and diagnosis of SCA2. They can also be used for the
CC treatment of SCA2 by antisense therapy or gene therapy.
SQ Sequence 203 BP; 68 A; 70 C; 62 G; 3 T;

Query Match 6.2%; Score 36.6; DB 1; Length 203;
Best Local Similarity 57.4%; Pred. No. 0.79;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 184 GCGGCGAGTAGCAGTCATACCAATACGACTTCTGCGGCTGCTGCTGATCCCTGCATCACCC 243
    || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 GCCCAGAGTTTCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143
    || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 244 CTTGTTTACGCCAGCCAGCAGCGCGCTCTCTGCGCGCGCGGCGATCGGCTG 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
```


John M;
PI DR WPI: 96-139095/14.
PT New isolated fibre-specific promoters - used for introducing
PT altered fibre-specific characteristics into plants, partic. cotton.
PS Claim 1; Column 59-62; 48pp; English.
CC The cotton var. Sea Island H6 gene and promoter (T13030) were
CC identified in genomic clone EMBL-S1-H6-4 isolated by screening a
CC genomic library prep. in lambda EMBL with cotton fibre-specific
CC cDNA clone H6 (see T13034). H6 and other (see T13025-29, T13031-32
CC and T13052-53) fibre-specific promoters can be utilised in the
CC construction of vectors used for the introduction of altered fibre-
CC specific characteristics into plants, partic. cotton. They can be
CC used to modulate the synthesis of fibre proteins or to introduce
CC non-fibre proteins into fibre in a tissue-specific manner.
SQ Sequence 1984 BP; 565 A; 470 C; 281 G; 668 T;

Query Match 6.2%; Score 36.6; DB 1; Length 1984;
Best Local Similarity 58.9%; Pred. No. 1.3;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 198 TCATACCAATAGACTTCTGCGCTGCTGTCATCCCTGTCATCCCTGTTTCAGCGCC 257
DB 569 TCCACCCAGTTCTCTCTCTGCACTCCACCACTGCTTCTCTCTCTGCAACTCC 628

QY 258 AGCCAAGGACGCGCTCTCTGCGCGCGCGGATCGGCTGAGCGCG 304
DB 629 ACCTCCAGCTTCTCCACTCTGCGCACTCTCTCCACAGCTTCTCCAC 675

RESULT 13
T30250
ID T30250 standard; DNA; 1985 BP.
AC T30250;
DE 16-DEC-1996 (first entry)
KW Cotton fibre clone SKSIE6-H6-RI-derived promoter contg. sequence.
KW Cotton fibre; promoter; differential screening; leaf; ovule; root;
KW flower; PCR; polymerase chain reaction; homology; transgenic plant; ds.
OS Gossypium barbadense.
FH Key Location/Qualifiers
FT Promoter
FT 1..250
FT /*tag= a
FT /note= "Contains promoter sequence (claimed)"
FT misc_feature 251..321
FT /*tag= b
FT /note= "5' non-coding sequence"
FT cds 322..1547
FT /*tag= c
FT /note= "Contains a 583 bp intron"
FT misc_feature 1548..1984
FT /*tag= d
FT /note= "3' non-coding sequence"
PN US521078-A.
PD 28-MAY-1996.
PF 04-OCT-1988; 253243.
PR 04-OCT-1988; US-253243.
PR 21-NOV-1990; US-617239.
PR 18-MAY-1992; US-885970.
PR 19-OCT-1994; US-298687.
PA (CETU) AGRACETUS INC.
PI John M;
DR WPI: 96-267794/27.
PT Isolation of fibre-specific cotton promoter sequences - using
PT selected DNA probes to screen genomic DNA fragments, for production
PT of cotton fibres with improved characteristics
PS Claim 1; Column 59-62; 46pp; English.
CC Cotton fibre cell-specific promoter sequences were isolated by
CC differential screening of a cotton plant cDNA library. Of 4788 clones
CC from a 10 day cell library screened with leaf cDNAs, 800 clones not
CC present in the leaf were isolated. These were screened with cDNAs from
CC ovule, root and flower mRNAs and resulted in 79 clones isolated. PCR
CC analysis was then used to remove cross-hybridising clones. This resulted
CC in the isolation of 18 cDNA clones specifically expressed in cotton fibre

cells (T30242-4 and T30253-67). These cDNAs were then used to screen for
CC homologous genomic sequences (T30245-53 and T30268) in order to obtain
CC the corresp. promoter sequences.
CC This genomic clone is derived from the 1.9 kb SalI-EcoRI fragment of the
CC sub-clone designated SKSIE6-H6-RI. This forms part of the genomic clone
CC designated as EMBL-S1-H6-4 which contains a 13 kb MboI insert.
CC The promoters isolated from the fibre cell-specific clones can be used to
CC generate transgenic cotton plants and lines producing fibres having
CC altered quantity and quality.
SQ Sequence 1985 BP; 565 A; 470 C; 282 G; 668 T;

Query Match 6.2%; Score 36.6; DB 1; Length 1985;
Best Local Similarity 58.9%; Pred. No. 1.3;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 198 TCATACCAATAGACTTCTGCGCTGCTGTCATCCCTGTCATCCCTGTTTCAGCGCC 257
DB 569 TCCACCCAGTTCTCTCTCTGCACTCCACCACTGCTTCTCTCTCTGCAACTCC 628

QY 258 AGCCAAGGACGCGCGCTCTCTGCGCGCGGATCGGCTGAGCGCG 304
DB 629 ACCTCCAGCTTCTCCACTCTGCGCACTCTCTCCACAGCTTCTCCAC 675

RESULT 14
T70036
ID T70036 standard; DNA; 1985 BP.
AC T70036;
DE 20-AUG-1997 (first entry)
KW Cotton H6 gene and fibre-specific promoter from clone SIH6.
KW cotton; E6; fibre; promoter; transgenic plant; truncated;
KW heterologous gene expression; ds.
OS Gossypium barbadense strain Sea Island.
PN US5620882-A.
PD 15-APR-1997.
PF 04-OCT-1988; 253243.
PR 04-OCT-1988; US-253243.
PR 21-NOV-1990; US-617239.
PR 18-MAY-1992; US-885970.
PR 19-OCT-1994; US-298829.
PA (CETU) AGRACETUS INC.
PI John M;
DR WPI: 97-235185/21.
PT DNA constructs contg. truncated promoter sequence - for
PT fibre-specific gene expression in cotton plants
PS Claim 1; Column 61-64; 48pp; English.
CC T70031-38 are genomic DNA clones containing cotton fibre-specific
CC promoters. Claimed DNA constructs comprise a truncated promoter sequence
CC (from one of T70031-38) that promotes preferential gene expression in
CC plant fibre cells, a protein coding sequence not naturally associated
CC with the promoter sequence and a 3' termination sequence. The DNA
CC constructs are useful for expressing foreign genes in fibre-producing
CC plants, esp. to produce transgenic cotton plants with varied cotton fibre
CC characteristics and quality. The present sequence comprises a fragment of
CC SIH6-4RI (an EcoRI fragment isolated from PSKSIH6-4 which hybridises to
CC H6 cDNA).
SQ Sequence 1985 BP; 565 A; 470 C; 282 G; 668 T;

Query Match 6.2%; Score 36.6; DB 1; Length 1985;
Best Local Similarity 58.9%; Pred. No. 1.3;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 198 TCATACCAATAGACTTCTGCGCTGCTGTCATCCCTGTCATCCCTGTTTCAGCGCC 257
DB 569 TCCACCCAGTTCTCTCTCTGCACTCCACCACTGCTTCTCTCTCTGCAACTCC 628

QY 258 AGCCAAGGACGCGCGCTCTCTGCGCGCGGATCGGCTGAGCGCG 304
DB 629 ACCTCCAGCTTCTCCACTCTGCGCACTCTCTCCACAGCTTCTCCAC 675

THIS PAGE BLANK (USPTO)